```
;      STRAIN:  NRRL 2338
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..6659
;      OTHER INFORMATION:  /function= "APPROXIMATE SPAN OF
;      OTHER INFORMATION:  MODULE 1"
;      OTHER INFORMATION:  /label= FUNCTION
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  744..11219
;      OTHER INFORMATION:  /function= "gene= "eryA""
;      OTHER INFORMATION:  /product= ""ORF1 encoding modules 1 & 2 for
;      OTHER INFORMATION:  6-deoxyerythronolide B""
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..1868
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1998..2198
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2250..3626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACP synthase domain/module1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3831..4811
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase 2 domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5574..6125
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6369..6626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 2 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..11219
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..8066
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACPsynthase of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  8262..9305
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9906..10454
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10707..10964
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 2"
US-08-439-009A-1

  Query Match             38.5%;  Score 94.6;  DB 3;  Length 11219;
  Best Local Similarity   61.6%;  Pred. No. 3.4e-17;
  Matches  151;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;
```

```
Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              |||||   || | ||   ||||| |    ||| |   || |    || || ||   |
Db      10707 CTGGCGGGTCTGCCGCGTGCCGAGCGGACGGCGGAGCTGGTGCGCCTGGTCCGCACCAGC 10766

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
                ||| | |||||  | ||||  |||||||| ||  |     |   |  | |||||
Db      10767 ACCGCGACCGTGCTGGGCCACGACGACCCGAAGGCGGTGCGCGCGACCACGCCGTTCAAG 10826

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |  ||||| ||||| || ||  ||||||||  | | || ||||||||| | |  ||| ||
Db      10827 GAGCTCGGGTTCGACTCGCTGGCGGCCGTCCGGCTGCGCAACCTGCTCAACGCGGCCACC 10886

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || |||||||| ||      |||||||||||||||||||| | |  ||    ||||||
Db      10887 GGGCTCCGCCTGCCGTCGACGCTGGTCTTCGACCACCCGAACGCCTCCGCGGTCGCCGGT 10946

Qy        241 CACCT 245
                |||
Db      10947 TTCCT 10951
```

Search completed: June 18, 2003, 01:07:08
Job time : 20.6565 secs

```
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Saccharopolyspora erythraea
;       STRAIN:  NRRL 2338
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  744..6659
;       OTHER INFORMATION:  /function= "APPROXIMATE SPAN OF
;       OTHER INFORMATION:  MODULE 1"
;       OTHER INFORMATION:  /label= FUNCTION
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  744..11219
;       OTHER INFORMATION:  /function= "gene= "eryA""
;       OTHER INFORMATION:  /product= ""ORF1 encoding modules 1 & 2 for
;       OTHER INFORMATION:  6-deoxyerythronolide B""
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  744..1868
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain 1 of module 1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1998..2198
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain 1 of module 1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  2250..3626
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACP synthase domain/module1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3831..4811
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase 2 domain of module 1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  5574..6125
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6369..6626
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain 2 of module 1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6678..11219
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 2"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6678..8066
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacyl ACPsynthase of module 2"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  8262..9305
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 2"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  9906..10454
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 2"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10707..10964
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 2"
US-07-642-734C-1

  Query Match             38.5%;  Score 94.6;  DB 1;  Length 11219;
  Best Local Similarity   61.6%;  Pred. No. 3.4e-17;
  Matches  151;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              ||||   || | ||   ||||| |    ||| |   || |   || || ||   |
Db      10707 CTGGCGGGTCTGCCGCGTGCCGAGCGGACGGCGGAGCTGGTGCGCCTGGTCCGCACCAGC 10766

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              ||| | |||||  | ||||  |||||||| ||  |     |  | | |||||
Db      10767 ACCGCGACCGTGCTGGGCCACGACGACCCGAAGGCGGTGCGCGCGACCACGCCGTTCAAG 10826

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              | ||||| ||||| || ||  |||||||  | | || |||||||| | |  ||| ||
Db      10827 GAGCTCGGGTTCGACTCGCTGGCGGCCGTCCGGCTGCGCAACCTGCTCAACGCGGCCACC 10886

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || |||||||| ||     ||||||||||||||||||||| | |  ||    ||||||
Db      10887 GGGCTCCGCCTGCCGTCGACGCTGGTCTTCGACCACCCGAACGCCTCCGCGGTCGCCGGT 10946

Qy        241 CACCT 245
              |||
Db      10947 TTCCT 10951


RESULT 15
US-08-439-009A-1
; Sequence 1, Application US/08439009A
; Patent No. 6004787
;  GENERAL INFORMATION:
;    APPLICANT:  Donadio, S
;    APPLICANT:  Katz, L
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Method of Directing Biosynthesis of
;    TITLE OF INVENTION:  Specific Polyketides
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steven F. Weinstock
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,009A
;      FILING DATE:  11-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Casuto, Dianne
;      REGISTRATION NUMBER:  40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847-938-3137
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11219 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
```

```
Qy          18 GGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCATGTGGCGGCAGTGCTCCG 77
               |||||  |  ||| | ||  | |||   ||||| ||   ||| | |||||| || ||| |
Db        4194 GCCCGAGCGCCGGCCGGCGCTCCTCACCCTCGTCCGTACCCACGCGGCGGCCGTACTCGG 4253

Qy          78 GCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGTGCACTCGGCTTCGATTC 137
               ||  |   |||  |     | || |||| ||   | |||   |  ||||||||||| ||
Db        4254 CCATTCCTCCCCCGACCGGGTGGCCCCCGGCCGTGCCTTCACCGAGCTCGGCTTCGACTC 4313

Qy         138 ACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACAGGACTCCGCCTTCCTGT 197
               || ||||||||  || |||| |||| |||   || ||     ||   | | || || |
Db        4314 GCTGACGGCCGTGCAGCTCCGCAACCAGCTCTCCACGGTGGTCGGCAACAGGCTCCCCGC 4373

Qy         198 CTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTACACCT 245
               | |  ||||||||||||||||||| || ||| ||||||||  |||||
Db        4374 CACCACGGTCTTCGACCACCCGACGCCCGCCGCACTCGCCGCGCACCT 4421


RESULT 12
US-09-434-288-3/c
; Sequence 3, Application US/09434288
; Patent No. 6303767
; GENERAL INFORMATION:
;  APPLICANT:  Betlach C., Melanie
;  APPLICANT:  McDaniel, Robert
;  TITLE OF INVENTION:  POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
;  TITLE OF INVENTION:  CONSTRUCTS THEREFOR
;  FILE REFERENCE:  30062-20030.00
;  CURRENT APPLICATION NUMBER: US/09/434,288
;  CURRENT FILING DATE:  1999-11-05
;  PRIOR APPLICATION NUMBER:  60/107,093
;  PRIOR FILING DATE:  1998-11-05
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 1434
;   TYPE: DNA
;   ORGANISM: Streptomyces narbonensis
US-09-434-288-3

  Query Match             39.8%;  Score 97.8;  DB 4;  Length 1434;
  Best Local Similarity   62.4%;  Pred. No. 3e-18;
  Matches  153;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy           1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
               ||||||     | |  | | ||| |    ||       | ||||  |||||||| | |||
Db        1161 CTGGCCGCCGCGGCTCCCGGCGAGCGTACGGAGATCCTCCTCGGTCTCGTACGGGCCCAG 1102

Qy          61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
               |  || || |||||||||        ||  | | | ||||  ||||||  | | |||
Db        1101 GCCGCCGCCGTGCTCCGGATGCGTTCACCTGAGGACGTCGCCGCCGACCGGGCCTTCAAG 1042

Qy         121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
               |  ||||||||||  || ||| | |  ||||||  |||| |||| |||||  |  ||| ||
Db        1041 GACATCGGCTTCGACTCGCTCGCCGGAGTCGAGCTCCGCAACCGGCTGACCCGGGCGACC 982

Qy         181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
               || |||| | | || |   |||| ||||||||||||||||||  ||        |||||   |
Db         981 GGGCTCCCCTTGCCCGCGACGCTCGTCTTCGACCACCCGACGCCGCTGGCCCTCGCGTCA 922

Qy         241 CACCT 245
               |  ||
Db         921 CTGCT 917


RESULT 13
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-1

  Query Match             39.1%;  Score 96.2;  DB 4;  Length 15872;
  Best Local Similarity   66.8%;  Pred. No. 1.3e-17;
  Matches  137;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          42 CGAGCTCGTACGAGACCATGTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGC 101
               |||||  ||| ||  |  || || || || || |||   | |||  |||||||     | |||
Db        2815 CGAGCGCGTGCGCGCACACGTCGCCGCCGTCCTCGAGTACGACGACCCGACCCGCGTCGA 2874

Qy         102 GCCCGACCAGTCGTTCCGTGCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAA 161
               | || ||   | |||   |  || |||||||| || |||| | |||||||| | || ||
Db        2875 ACTCGGCCTCACCTTCAAGGAGCTGGGCTTCGACTCCCTCATGTCCGTCGAGCTGCGGAA 2934

Qy         162 CCTGCTGATCAAGGCAACAGGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGAC 221
               |  |||  || | |  || ||||| |||||  ||    |  ||| ||||||||||||||||
Db        2935 CGCGCTCGTCGACGACACGGGACTGCGCCTGCCCAGCGGACTGCTCTTCGACCACCCGAC 2994

Qy         222 CCCTGCCAAACTCGCCGTACACCTG 246
               ||   |  |||||||  ||||||
Db        2995 GCCGCGCGCCCTCGCCGCCCACCTG 3019


RESULT 14
US-07-642-734C-1
; Sequence 1, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, L
;    APPLICANT:  Donadio, S
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Recombinant DNA Method for Producing
;    TITLE OF INVENTION:  Erythromycin Analogs
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Danckers, Andreas M
;      REGISTRATION NUMBER:  32652
;      REFERENCE/DOCKET NUMBER:  4952.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11219 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
```

```
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  80161 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-036-987A-1

  Query Match             40.8%;  Score 100.4;  DB 3;  Length 80161;
  Best Local Similarity   63.0%;  Pred. No. 1.3e-18;
  Matches  155;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || ||  | | | ||   ||| | | |||   || | || | || || ||  | | |||
Db      39790 CTCGCGGAGCTGCCGGAAGCCCAACGACGCCACGAACTGGTGGATCTGGTGTGCGCCCAG 39849

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |||||  | |||||| ||||||     |   | |   ||  |||||| | | |||||||
Db      39850 GTGGCAACCGTGCTCGGGCACGGCAGTCGCGAGGAAGTCCAGCCCGAGCGGGCGTTCCGC 39909

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              || ||||| ||||| || |||| ||| || ||  | || || |   ||| ||  || ||
Db      39910 GCGCTCGGGTTCGACTCCCTCATGGCGGTGGATCTGCGCAATCGTTTGACCACCGCCACC 39969

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              ||  | ||||| ||   | |    |||||||| |||||  || |||    | ||||
Db      39970 GGGTTGCGCCTGCCGACCACAACCGTCTTCGACTACCCGAATCCGGCCGCCTTGGCCGCT 40029

Qy        241 CACCTG 246
              ||||||
Db      40030 CACCTG 40035


RESULT 10
US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 80161
;   TYPE: DNA
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

  Query Match             40.8%;  Score 100.4;  DB 4;  Length 80161;
  Best Local Similarity   63.0%;  Pred. No. 1.3e-18;
  Matches  155;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || ||  | | | ||   ||| | | |||   || | || | || || ||  | | |||
Db      39790 CTCGCGGAGCTGCCGGAAGCCCAACGACGCCACGAACTGGTGGATCTGGTGTGCGCCCAG 39849

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |||||  | |||||| ||||||     |   | |   ||  |||||| | | |||||||
Db      39850 GTGGCAACCGTGCTCGGGCACGGCAGTCGCGAGGAAGTCCAGCCCGAGCGGGCGTTCCGC 39909

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              || ||||| ||||| || |||| ||| || ||  | || || |   ||| ||  || ||
Db      39910 GCGCTCGGGTTCGACTCCCTCATGGCGGTGGATCTGCGCAATCGTTTGACCACCGCCACC 39969

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              ||  | ||||| ||   | |    |||||||| |||||  || |||    | ||||
Db      39970 GGGTTGCGCCTGCCGACCACAACCGTCTTCGACTACCCGAATCCGGCCGCCTTGGCCGCT 40029

Qy        241 CACCTG 246
              ||||||
Db      40030 CACCTG 40035


RESULT 11
US-09-105-537-34
; Sequence 34, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-34

  Query Match             40.0%;  Score 98.4;  DB 4;  Length 4689;
  Best Local Similarity   64.5%;  Pred. No. 2.6e-18;
  Matches  147;  Conservative    0;  Mismatches   81;  Indels    0;  Gaps    0;
```

```
;       CITY:  INDIANAPOLIS
;       STATE:  IN
;       COUNTRY:  USA
;       ZIP:  46285
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  MS-DOS
;       SOFTWARE:  ASCI(DOS) Text only
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/804,227C
;       FILING DATE:  February 21, 1997
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Plant, Thomas, G.
;       REGISTRATION NUMBER:  35,784
;       REFERENCE/DOCKET NUMBER:  X-8231
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  317-276-2459
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  44377 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  350..14002
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  14046..20036
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  20110..31284
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  31329..36071
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  36155..41830
US-08-804-227C-7

  Query Match             40.9%;  Score 100.6;  DB 2;  Length 44377;
  Best Local Similarity   65.2%;  Pred. No. 1e-18;
  Matches  148;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy          20 CCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCATGTGGCGGCAGTGCTCCGGC 79
               |||| |    |||     ||||||| || || ||    | | || ||||| ||||||  | |
Db       19495 CCGAGCGTACGGCTCTCCTGCTCGACCTGGTGCGGACCGAGGTCGCGGCGGTGCTCGGAC 19554

Qy          80 ACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGTGCACTCGGCTTCGATTCAC 139
               |||  |||||    || ||||   ||| ||   | |||    |    ||| ||||| || |
Db       19555 ACGGCGACCCCGCCGCGATCGGCGCCGCCCGCACCTTCAAGGACGCCGGATTCGACTCCC 19614

Qy         140 TCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACAGGACTCCGCCTTCCTGTCT 199
               |||| || |||||  |||| |||| ||||| ||     || ||||| || || || |
Db       19615 TCACCGCTGTCGACCTCCGCAACCGGCTGAACACACGCACCGGACTGCGGCTGCCCGCGA 19674

Qy         200 CGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTACACCTG 246
               | || ||||||||||||||| || ||    |   ||||||| || ||||
Db       19675 CCCTCGTCTTCGACCACCCCACACCGCTCGCCCTCGCCGAACTCCTG 19721


RESULT 8
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-198-1

  Query Match             40.9%;  Score 100.6;  DB 2;  Length 44377;
  Best Local Similarity   65.2%;  Pred. No. 1e-18;
  Matches  148;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy          20 CCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCATGTGGCGGCAGTGCTCCGGC 79
               |||| |    |||     ||||||| || || ||    | | || ||||| ||||||  | |
Db       19495 CCGAGCGTACGGCTCTCCTGCTCGACCTGGTGCGGACCGAGGTCGCGGCGGTGCTCGGAC 19554

Qy          80 ACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGTGCACTCGGCTTCGATTCAC 139
               |||  |||||    || ||||   ||| ||   | |||    |    ||| ||||| || |
Db       19555 ACGGCGACCCCGCCGCGATCGGCGCCGCCCGCACCTTCAAGGACGCCGGATTCGACTCCC 19614

Qy         140 TCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACAGGACTCCGCCTTCCTGTCT 199
               |||| || |||||  |||| |||| ||||| ||     || ||||| || || || |
Db       19615 TCACCGCTGTCGACCTCCGCAACCGGCTGAACACACGCACCGGACTGCGGCTGCCCGCGA 19674

Qy         200 CGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTACACCTG 246
               | || ||||||||||||||| || ||    |   ||||||| || ||||
Db       19675 CCCTCGTCTTCGACCACCCCACACCGCTCGCCCTCGCCGAACTCCTG 19721


RESULT 9
US-09-036-987A-1
; Sequence 1, Application US/09036987A
```

```
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-19

  Query Match             42.1%;  Score 103.6;  DB 3;  Length 38506;
  Best Local Similarity   63.8%;  Pred. No. 1.5e-19;
  Matches  157;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||     |||  ||| |||||     ||     |||  || ||||| ||     ||
Db      24507 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 24566

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |||||| | || ||  | ||||    |||||      | |   |  || | | | |||||
Db      24567 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 24626

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |    ||| ||||| || ||||| ||||||||  |||| ||||  || | |   || ||
Db      24627 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 24686

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || | | ||||||||||||||||||| ||||| |   | |||||||
Db      24687 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 24746

Qy        241 CACCTG 246
              ||||||
Db      24747 CACCTG 24752


RESULT 6
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  43280 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  816..14234
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14351..19945
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20010..31199
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31232..36067
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36249..41774
US-08-804-227C-1

  Query Match             41.5%;  Score 102;  DB 2;  Length 43280;
  Best Local Similarity   63.4%;  Pred. No. 4.2e-19;
  Matches  156;  Conservative    0;  Mismatches   90;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              ||| ||   |  ||| |   ||   |  |||   | || |      | || | |||
Db      30648 CTGACCGGGCTCACGGCACAGGAAGGGCACGCCCTCGTCCTGGCCGAGATCCGCGCCCAC 30707

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              | |||||| |||||  | ||||     |    | | ||| |     |||| | | |||
Db      30708 GCGGCGGCGGTGCTGGGCCACGGCTCCGACGACTCGATCCCCGAGGACCGGGCCTTCAAG 30767

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |  |||||||||| || ||||| ||||| ||| | || |||| |||||  |  ||| ||
Db      30768 GACCTCGGCTTCGACTCGCTCACCGCCGTGGAGATGCGCAACCGGCTGAGCGCGGCCACC 30827

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || ||||| || || | | | || |||||||||||||||||||| | | | || ||||
Db      30828 GGCCTCCGGCTGCCCGCCACCCTCGTCTTCGACCACCCGACCCCGGGCGAGCTGGCCGGC 30887

Qy        241 CACCTG 246
              ||||||
Db      30888 CACCTG 30893


RESULT 7
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
```

```
; Sequence 7, Application US/09434288
; Patent No. 6303767
; GENERAL INFORMATION:
;  APPLICANT:  Betlach C., Melanie
;  APPLICANT:  McDaniel, Robert
;  TITLE OF INVENTION:  POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
;  TITLE OF INVENTION:  CONSTRUCTS THEREFOR
;  FILE REFERENCE:  30062-20030.00
;  CURRENT APPLICATION NUMBER: US/09/434,288
;  CURRENT FILING DATE:  1999-11-05
;  PRIOR APPLICATION NUMBER:  60/107,093
;  PRIOR FILING DATE:  1998-11-05
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 7
;   LENGTH: 1681
;   TYPE: DNA
;   ORGANISM: Streptomyces narbonensis
US-09-434-288-7

  Query Match             42.1%;  Score 103.6;  DB 4;  Length 1681;
  Best Local Similarity   63.8%;  Pred. No. 8e-20;
  Matches  157;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||      |||  ||| ||||       ||      |||  || ||||| ||     ||
Db        541 CTCGCCGCGATGACCCCGGACGACAGGACCGCGCACCTGCGGGACCTCGTCCGTACGCAC 600

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |||||| | || ||  | ||||  | |||||       |||  |  || |    | |||||
Db        601 GTGGCGACCGTCCTGGGACACGGCGCCCCGAGCCGGGTCGACCTGGAGCGCGCCTTCCGC 660

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |    ||| ||||| || |||||| |||||||| |||| ||||  || | |   || ||
Db        661 GACACCGGTTTCGACTCCCTCACCGCCGTCGAGCTCCGCAACCGCCTCAACGCCGCCACC 720

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || ||||| || | | |||| ||||||||||||||| || || |    | |||||||
Db        721 GGGCTGCGCCTCCCGGCCACGCTCGTCTTCGACCACCCCACTCCGGGGGAGCTCGCCGGG 780

Qy        241 CACCTG 246
              ||||||
Db        781 CACCTG 786


RESULT 3
US-09-105-537-32
; Sequence 32, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-32

  Query Match             42.1%;  Score 103.6;  DB 4;  Length 11220;
  Best Local Similarity   63.8%;  Pred. No. 1.2e-19;
  Matches  157;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||      |||  ||| |||||      ||      |||  || ||||| ||     ||
Db      10678 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 10737

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |||||| | || ||  | ||||  |||||       |  |  || | | | |||||
Db      10738 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 10797

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |    ||| ||||| || ||||| ||||||||  |||| ||||  || | |   || ||
Db      10798 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 10857

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || | | ||||||||||||||||||| ||||| |   | |||||||
Db      10858 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 10917

Qy        241 CACCTG 246
              ||||||
Db      10918 CACCTG 10923


RESULT 4
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-5

  Query Match             42.1%;  Score 103.6;  DB 4;  Length 36778;
  Best Local Similarity   63.8%;  Pred. No. 1.5e-19;
  Matches  157;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||      |||  ||| |||||      ||      |||  || ||||| ||     ||
Db      26365 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 26424

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |||||| | || ||  | ||||  |||||       |  |  || | | | |||||
Db      26425 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 26484

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |    ||| ||||| || ||||| ||||||||  |||| ||||  || | |   || ||
Db      26485 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 26544

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || | | ||||||||||||||||||| ||||| |   | |||||||
Db      26545 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 26604

Qy        241 CACCTG 246
              ||||||
Db      26605 CACCTG 26610


RESULT 5
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 10.6565 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_5935_6180
Perfect score: 246
Sequence: 1 ctggcccatcagacgtcggc.....ccaaactcgccgtacacctg 246

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	110.2	44.8	50937	4	US-09-428-517-1	Sequence 1, Appli
	2	103.6	42.1	1681	4	US-09-434-288-7	Sequence 7, Appli
	3	103.6	42.1	11220	4	US-09-105-537-32	Sequence 32, Appl
	4	103.6	42.1	36778	4	US-09-105-537-5	Sequence 5, Appli
	5	103.6	42.1	38506	3	US-09-320-878-19	Sequence 19, Appl
	6	102	41.5	43280	2	US-08-804-227C-1	Sequence 1, Appli
	7	100.6	40.9	44377	2	US-08-804-227C-7	Sequence 7, Appli
	8	100.6	40.9	44377	2	US-08-804-198-1	Sequence 1, Appli
	9	100.4	40.8	80161	3	US-09-036-987A-1	Sequence 1, Appli
	10	100.4	40.8	80161	4	US-09-370-700-1	Sequence 1, Appli
	11	98.4	40.0	4689	4	US-09-105-537-34	Sequence 34, Appl
c	12	97.8	39.8	1434	4	US-09-434-288-3	Sequence 3, Appli
	13	96.2	39.1	15872	4	US-09-105-537-1	Sequence 1, Appli
	14	94.6	38.5	11219	1	US-07-642-734C-1	Sequence 1, Appli
	15	94.6	38.5	11219	3	US-08-439-009A-1	Sequence 1, Appli
	16	94.6	38.5	13842	4	US-09-105-537-30	Sequence 30, Appl
c	17	92.8	37.7	561	3	US-09-154-083-13	Sequence 13, Appl
	18	91	37.0	20235	1	US-07-642-734C-3	Sequence 3, Appli
	19	91	37.0	20235	3	US-08-439-009A-3	Sequence 3, Appli
	20	89	36.2	13987	2	US-08-804-227C-13	Sequence 13, Appl
	21	86	35.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	22	86	35.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	23	84.4	34.3	28958	1	US-08-258-261B-6	Sequence 6, Appli
	24	84.4	34.3	28958	1	US-08-456-837-6	Sequence 6, Appli
	25	84.4	34.3	28958	1	US-08-457-342-6	Sequence 6, Appli
	26	84.4	34.3	28958	1	US-08-457-646A-6	Sequence 6, Appli
	27	84.4	34.3	28958	1	US-08-458-076A-6	Sequence 6, Appli
	28	84.4	34.3	28958	1	US-08-764-233A-4	Sequence 4, Appli
	29	84.4	34.3	28958	1	US-08-457-335A-6	Sequence 6, Appli
	30	84.4	34.3	28958	1	US-08-729-214-6	Sequence 6, Appli
	31	84.4	34.3	28958	3	US-09-028-934-6	Sequence 6, Appli
	32	84.4	34.3	49377	1	US-08-764-233A-1	Sequence 1, Appli
c	33	84.4	34.3	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	34	84.4	34.3	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	35	82.8	33.7	367	4	US-09-060-756-627	Sequence 627, App
c	36	82.6	33.6	444	4	US-09-060-756-658	Sequence 658, App
	37	80.8	32.8	4041	4	US-09-105-537-36	Sequence 36, Appl
	38	78.6	32.0	33529	4	US-09-144-085-3	Sequence 3, Appli
c	39	65.6	26.7	748	3	US-09-154-083-1	Sequence 1, Appli
c	40	65.6	26.7	753	3	US-09-154-083-28	Sequence 28, Appl
	41	59	24.0	68750	3	US-09-335-409-1	Sequence 1, Appli
	42	59	24.0	68750	4	US-09-568-102-1	Sequence 1, Appli
	43	59	24.0	68750	4	US-09-567-969-1	Sequence 1, Appli
	44	59	24.0	68750	4	US-09-568-480-1	Sequence 1, Appli
	45	59	24.0	68750	4	US-09-568-486-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artifical Sequence: Recombinant DNA
US-09-428-517-1

Query Match 44.8%; Score 110.2; DB 4; Length 50937;
Best Local Similarity 66.8%; Pred. No. 2.5e-21;
Matches 157; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```
Qy         12 GACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCATGTGGCGGCAGT 71
              | || |  | |||| |||     ||  ||| ||| ||||| || | ||| || ||    ||
Db      29150 GGCGGCACCGGACCCACGTCGTGCCCTGCGCGACCTCGTCCGCGGCCACGTCGCCCTGGT 29209

Qy         72 GCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGTGCACTCGGCTT 131
              |||| | |||  | |||  | |||||||   |||| ||| |.|||| |  |||| ||
Db      29210 GCTCGGACACAGTGGCCCCGAGGCCATCGACGCCGAACAGGCCTTCCGGGACATCGGTTT 29269

Qy        132 CGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACAGGACTCCGCCT 191
              ||| || || || || |||||  || ||||||  ||||| |   |  || || |||||| |
Db      29270 CGACTCCCTGACCGCAGTCGAACTCAGAAACCGGCTGAACGCCGAGACCGGCCTCCGCTT 29329

Qy        192 TCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTACACCTG 246
              || | | |||| || |||||| |||| | |||  |  |||||||  ||||||
Db      29330 GCCCGGCACGCTCGTGTTCGACTACCCCAACCCGAGCGCGCTCGCCGATCACCTG 29384
```

RESULT 2
US-09-434-288-7

THIS PAGE BLANK (USPTO)

PR 06-MAY-1998; 98US-0073538.
PR 28-AUG-1998; 98US-0141908.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Ashley G, Betlach MC, Betlach M, Tang L, McDaniel R;
XX
DR WPI; 2000-610844/58.
XX
PT New recombinant picK hydroxylase gene of Streptomyces venezuelae useful
PT for converting ketolides to antibiotics and as antibiotics and
PT intermediates in the synthesis of compounds with pharmaceutical value
PT -
XX
PS Disclosure; Columns 15-32; 117pp; English.
XX
CC The present sequence is used to produce the recombinant DNA compounds
CC of the invention. The specification describes a recombinant DNA compound
CC expressing recombinant polyketide synthase genes in host cells for the
CC production of narbonolide, narbonolide derivatives and polyketides that
CC are useful as antibiotics and as intermediates in the synthesis of
CC compounds with pharmaceutical value. The DNA compounds may also encode
CC a C12-hydroxyalse (picK), desosamine biosynthesis and desosaminyl
CC transferase enzymes (useful for conversion of ketolides to antibiotics),
CC and the beta-glucosidase enzyme (involved in picromycin biosynthesis).
CC These compounds are also useful for increasing the antibiotic activity
CC of a compound relative to the unhydroxylated compound. The recombinant
CC host cells are useful as genetic systems that allow rapid engineering
CC of the narbonolide polyketide synthase. These would be valuable for
CC creating novel ketolide analogs for pharmaceutical applications.
XX
SQ Sequence 38506 BP; 4914 A; 15118 C; 13444 G; 5030 T; 0 other;

Query Match 42.1%; Score 103.6; DB 21; Length 38506;
Best Local Similarity 63.8%; Pred. No. 2.5e-18;
Matches 157; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

```
Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||     |||  ||| |||||     ||     |||  || ||||| ||    ||
Db      24507 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 24566

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |||||| | || ||  | ||||    |||||       | |  |  || | | | |||||
Db      24567 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 24626

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |    ||| ||||| || ||||| |||||||| |||| ||||  || | |   || ||
Db      24627 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 24686

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || | | ||||||||||||||||||| ||||| |   | |||||||
Db      24687 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 24746

Qy        241 CACCTG 246
              ||||||
Db      24747 CACCTG 24752
```

Search completed: June 17, 2003, 13:48:30
Job time : 51.0269 secs

```
  Query Match             42.1%;  Score 103.6;  DB 21;  Length 36778;
  Best Local Similarity   63.8%;  Pred. No. 2.5e-18;
  Matches  157;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||     |||  ||| |||||     ||     |||  || ||||| ||    ||
Db      26365 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 26424

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |||||| | || ||  | ||||    |||||       | |  |  || | | | |||||
Db      26425 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 26484

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |    ||| ||||| || ||||| ||||||||  |||| ||||  || | |   || ||
Db      26485 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 26544

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || | | ||||||||||||||||||| ||||| |   | |||||||
Db      26545 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 26604

Qy        241 CACCTG 246
              ||||||
Db      26605 CACCTG 26610


RESULT 14
AAZ87285
ID   AAZ87285 standard; DNA; 37948 BP.
XX
AC   AAZ87285;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae pik (macrolide biosynthesis) gene cluster, SEQ ID NO:5.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77180.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 13; Page 299-315; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the macrolide
CC   biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC 15439.
XX
SQ   Sequence 37948 BP; 4901 A; 14718 C; 13311 G; 5018 T; 0 other;

  Query Match             42.1%;  Score 103.6;  DB 21;  Length 37948;
  Best Local Similarity   63.8%;  Pred. No. 2.5e-18;
  Matches  157;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||     |||  ||| |||||     ||     |||  || ||||| ||    ||
Db      27535 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 27594

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |||||| | || ||  | ||||    |||||       | |  |  || | | | |||||
Db      27595 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 27654

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |    ||| ||||| || ||||| ||||||||  |||| ||||  || | |   || ||
Db      27655 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 27714

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || | | ||||||||||||||||||| ||||| |   | |||||||
Db      27715 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 27774

Qy        241 CACCTG 246
              ||||||
Db      27775 CACCTG 27780


RESULT 15
AAA75633
ID   AAA75633 standard; DNA; 38506 BP.
XX
AC   AAA75633;
XX
DT   22-JAN-2001  (first entry)
XX
DE   Nucleotide sequence of the insert DNA in cosmid pKOS023-27.
XX
KW   Narbonolide synthase; polyketide synthase gene; narbonolide polyketide;
KW   antibiotic; C12-hydroxyalse; picK; desosamine biosynthesis;
KW   desosaminyl transferase enzyme; ketolide; beta-glucosidase enzyme;
KW   picromycin biosynthesis; ss.
XX
OS   Streptomyces venezuelae.
XX
PN   US6117659-A.
XX
PD   12-SEP-2000.
XX
PF   27-MAY-1999;   99US-0320878.
XX
PR   28-MAY-1998;   98US-0087080.
PR   22-SEP-1998;   98US-0100880.
PR   08-FEB-1999;   99US-0119139.
PR   20-MAY-1999;   99US-0134990.
PR   30-APR-1997;   97US-0846247.
```

```
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Claim 15; Page 398-403; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the
CC   production of biologically active macrolides. The macrolide biosynthetic
CC   proteins are useful for synthesis of methymycin, pikromycin,
CC   neomethymycin and narbomycin. The alternative termination of polyketide
CC   synthesis may be useful to prepare novel antibiotics and
CC   polyhydroxyalkanoate (PHA) monomers. The compounds produced by the
CC   recombinant host cells are useful as biopolymers, e.g., in packaging or
CC   biomedical applications, to engineer PHA monomer synthases or to prepare
CC   biologically active agents, such as chemotherapeutics,
CC   immunosuppressants, agents to treat asthma, chronic obstructive pulmonary
CC   disease as well as other diseases involving respiratory inflammation,
CC   cholesterol-lowering agents or macrolide-based antibiotics which are
CC   active against a variety of organisms, e.g., bacteria, including
CC   multi-drug resistant pneumococci and other respiratory pathogens, as well
CC   as viral parasitic pathogens, or as crop protection agents (e.g.,
CC   fungicides or insecticides) via expression of polyketides in plants.
CC   Sequences AAZ87295-Z87302 represent macrolide biosynthetic genes from
CC   Streptomyces venezuelae ATCC 15439, which encode proteins
CC   AAY77190-Y77197.
XX
SQ   Sequence 11220 BP; 1369 A; 4423 C; 3966 G; 1462 T; 0 other;

  Query Match             42.1%;  Score 103.6;  DB 21;  Length 11220;
  Best Local Similarity   63.8%;  Pred. No. 2.1e-18;
  Matches  157;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||     |||  ||| |||||     ||     |||  || ||||| ||    ||
Db      10678 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 10737

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |||||| | || ||  | ||||    |||||      | |  |  || | | | |||||
Db      10738 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 10797

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |   ||| ||||| || ||||| ||||||||  |||| ||||  || | |    || ||
Db      10798 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 10857

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || | | ||||||||||||||||||| ||||| |   | |||||||
Db      10858 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 10917

Qy        241 CACCTG 246
              ||||||
Db      10918 CACCTG 10923


RESULT 13
AAZ87318
ID   AAZ87318 standard; DNA; 36778 BP.
XX
AC   AAZ87318;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae pik (macrolide biosynthesis) gene cluster.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             1742..15583
FT                   /*tag= a
FT                   /product= "Pik gene cluster protein #1 (AAY77200)"
FT   CDS             15688..26907
FT                   /*tag= b
FT                   /product= "Pik gene cluster protein #2 (AAY77201)"
FT   CDS             26991..31679
FT                   /*tag= c
FT                   /product= "Pik gene cluster protein #3 (AAY77202)"
FT   CDS             31782..35822
FT                   /*tag= d
FT                   /product= "Pik gene cluster protein #4 (AAY77203)"
FT   CDS             35819..36664
FT                   /*tag= e
FT                   /product= "Pik gene cluster protein #4 (AAY80997)"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77200, AAY77201, AAY77202, AAY77203, AAY80997.
XX
PT   Desosamine and macrolide biosynthetic gene clusters, useful for, e .g.
PT   synthesis of methymycin and pikromycin -
XX
PS   Disclosure; Figure 31; 438pp; English.
XX
CC   The invention relates to an isolated and purified nucleic acid segment
CC   comprising a desosamine biosynthetic gene cluster, a fragment or its
CC   biologically active variant, where the nucleic acid sequence is not
CC   derived from the eryC gene cluster of Saccharopolyspora erythraea or
CC   Streptomyces antibioticus. The invention also relates to a macrolide
CC   biosynthetic gene cluster, or fragments thereof. The macrolide
CC   biosynthetic gene cluster encodes proteins which synthesise methymycin,
CC   pikromycin, neomethymycin, narbomycin or a combination of these
CC   compounds. Recombinant or augmented cells comprising the desosamine
CC   and/or macrolide biosynthetic gene clusters are useful for the production
CC   of biologically active macrolides. The macrolide biosynthetic proteins
CC   are useful for synthesis of methymycin, pikromycin, neomethymycin and
CC   narbomycin. The alternative termination of polyketide synthesis may be
CC   useful to prepare novel antibiotics and polyhydroxyalkanoate (PHA)
CC   monomers. The compounds produced by the recombinant host cells are useful
CC   as biopolymers, e.g., in packaging or biomedical applications, to
CC   engineer PHA monomer synthases or to prepare biologically active agents,
CC   such as chemotherapeutics, immunosuppressants, agents to treat asthma,
CC   chronic obstructive pulmonary disease as well as other diseases involving
CC   respiratory inflammation, cholesterol-lowering agents or macrolide-based
CC   antibiotics which are active against a variety of organisms, e.g.,
CC   bacteria, including multi-drug resistant pneumococci and other
CC   respiratory pathogens, as well as viral parasitic pathogens, or as crop
CC   protection agents (e.g., fungicides or insecticides) via expression of
CC   polyketides in plants. The present sequence represents the macrolide
CC   biosynthetic gene cluster (pik) from Streptomyces venezuelae ATCC
CC   15439, as given in figure 31.
XX
SQ   Sequence 36778 BP; 4758 A; 14303 C; 12864 G; 4851 T; 2 other;
```

```
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nysl DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

  Query Match            42.8%;  Score 105.2;  DB 22;  Length 65140;
  Best Local Similarity  64.2%;  Pred. No. 1e-18;
  Matches  158;  Conservative    0;  Mismatches   88;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||    |       | |||| ||  ||||  ||  | || || ||||| ||   |||
Db      21497 CTCGCCGGCCTCGACGCCGCCGCCCGCCGGGAAGCCCTCCTGGAACTCGTCCGCACCCAG 21556

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
               | ||    || ||| | |||||||| | |  |   | | | ||| ||    |||||
Db      21557 ATCGCCCAGGTCCTCGGCCACGCGGACGCCACCGAGGTGGAGACCGGCCGCCAGTTCCAG 21616

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |  |||||||||| || ||||| ||||||||  |||| |||  |||| ||  || ||
Db      21617 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTCCGCAACGCCCTGAACACCGCCACC 21676

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || | | |  |||| |||||| ||||||| ||    |  |||||||
Db      21677 GGCCTGCGGCTGCCCGCCACCATGGTGTTCGACTACCCGACACCACACGCCCTCGCCGAC 21736

Qy        241 CACCTG 246
              ||||||
Db      21737 CACCTG 21742


RESULT 11
AAS18438
ID   AAS18438 standard; DNA; 1681 BP.
XX
AC   AAS18438;
XX
DT   12-MAR-2002  (first entry)
XX
DE   Contig 115 DNA encoding S. narbonensis polyketide synthase.
XX
KW   Narbonolide polyketide synthase; PKS; narbomycin modification enzyme;
KW   erythromycin; rapamycin; tylosin; picromycin; methylmycin; neomethymycin;
KW   agriculture; ds.
XX
OS   Streptomyces narbonensis.
XX
PN   US6303767-B1.
XX
PD   16-OCT-2001.
XX
PF   05-NOV-1999;   99US-0434288.
XX
PR   05-NOV-1998;   98US-107093P.
PR   27-MAY-1999;   99US-0320878.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Betlach MC,  McDaniel R;
XX
DR   WPI; 2002-065495/09.
XX
PT   Nucleic acids encoding narbonolide polyketide synthases from
PT   Streptomyces narbonensis, useful for the recombinant production of
PT   polyketides, e.g. narbomycin -
XX
PS   Claim 1; Column 18; 24pp; English.
XX
CC   The present invention relates to recombinant DNA vectors (cosmids)
CC   that encode for the narbonolide polyketide synthase (PKS) enzyme and
CC   various narbomycin modification enzymes from Streptomyces narbonensis.
CC   The recombinant DNA vectors can be used to produce recombinant ketide
CC   synthases and a variety of different polyketides (e.g. erythromycin,
CC   rapamycin, tylosin, narbomycin, picromycin, methylmycin and
CC   neomethymycin) for use in agriculture, medicine and health. The
CC   recombinant vectors may be used to produce polyketides in relatively
CC   high yields. AAS18432-AAS18443 represent contig DNA sequences that
CC   encode for S. narbonensis PKS enzymes.
XX
SQ   Sequence 1681 BP; 231 A; 672 C; 580 G; 198 T; 0 other;

  Query Match            42.1%;  Score 103.6;  DB 24;  Length 1681;
  Best Local Similarity  63.8%;  Pred. No. 1.5e-18;
  Matches  157;  Conservative    0;  Mismatches   89;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||     |||  ||| ||||      ||     |||  || ||||| ||    ||
Db        541 CTCGCCGCGATGACCCCGGACGACAGGACCGCGCACCTGCGGGACCTCGTCCGTACGCAC 600

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |||||| | || ||  | ||||  | |||||      |||  |  || |   | |||||
Db        601 GTGGCGACCGTCCTGGGACACGGCGCCCCGAGCCGGGTCGACCTGGAGCGCGCCTTCCGC 660

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |    ||| ||||| || ||||| |||||||| |||| ||||  || | |   || ||
Db        661 GACACCGGTTTCGACTCCCTCACCGCCGTCGAGCTCCGCAACCGCCTCAACGCCGCCACC 720

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || ||||| || | | |||| ||||||||||||| || || |   | |||||||
Db        721 GGGCTGCGCCTCCCGGCCACGCTCGTCTTCGACCACCCCACTCCGGGGGAGCTCGCCGGG 780

Qy        241 CACCTG 246
              ||||||
Db        781 CACCTG 786


RESULT 12
AAZ87298
ID   AAZ87298 standard; DNA; 11220 BP.
XX
AC   AAZ87298;
XX
DT   05-JUN-2000  (first entry)
XX
DE   S. venezuelae macrolide biosynthetic gene pikAII, SEQ ID NO:32.
XX
KW   Desosamine biosynthesis; macrolide; polyketide; methymycin; pikromycin;
KW   neomethymycin; narbomycin; polyhydroxyalkanoate monomer synthase;
KW   biopolymer; antibiotic; chemotherapeutic; immunosuppressant; asthma;
KW   chronic obstructive pulmonary disease; respiratory inflammation;
KW   hypercholesterolaemia; crop protection agent; ds.
XX
OS   Streptomyces venezuelae ATCC15439.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..111220
FT                   /*tag=  a
FT                   /product= "PikAII"
XX
PN   WO200000620-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US14398.
XX
PR   26-JUN-1998;   98US-0105537.
XX
PA   (MINU ) UNIV MINNESOTA.
XX
PI   Sherman DH,  Liu H,  Xue Y,  Zhao L;
XX
DR   WPI; 2000-160679/14.
DR   P-PSDB; AAY77193.
```

```
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65266, AAG65267.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Disclosure; Page 103-149; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;

  Query Match             43.7%;  Score 107.4;  DB 22;  Length 31422;
  Best Local Similarity   64.9%;  Pred. No. 2.3e-19;
  Matches  159;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              ||||||   |||||    |   | || |   || |  | ||||  || || ||   |||
Db      20089 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 20148

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
               | ||  | || ||| | ||| |   |||  |  ||||| | |||||||   |||||||
Db      20149 ATCGCCACCGTCCTCGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 20208

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              | ||||||||||| || ||||| ||||||||  | || |||| |||   |    | ||
Db      20209 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 20268

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              |||||||||| ||   | | || | |||||||||||| | |||  |||  ||| ||
Db      20269 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 20328

Qy        241 CACCT 245
              |||||
Db      20329 CACCT 20333


RESULT 10
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..1035)
FT                   /*tag=  a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement (1056..2576)
FT                   /*tag=  b
FT                   /product= "NysD1 protein"
FT   CDS             2806..6906
FT                   /*tag=  c
FT                   /product= "NysA protein"
FT   CDS             6952..16530
FT                   /*tag=  d
FT                   /product= "NysB protein"
FT   CDS             16550..49840
FT                   /*tag=  e
FT                   /product= "NysC protein"
FT   CDS             50260..51015
FT                   /*tag=  f
FT                   /product= "NysE protein"
FT   CDS             51405..54305
FT                   /*tag=  g
FT                   /product= "NysR1 protein"
FT   CDS             54329..57190
FT                   /*tag=  h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180..59963
FT                   /*tag=  i
FT                   /product= "NysR3 protein"
FT   CDS             60415..61047
FT                   /*tag=  j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736..62497
FT                   /*tag=  k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (62551..63615)
FT                   /*tag=  l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765..64961
FT                   /*tag=  m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 116-151; 266pp; English.
```

```
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match             44.7%;  Score 110;  DB 22;  Length 125401;
  Best Local Similarity   65.4%;  Pred. No. 5.8e-20;
  Matches  161;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || ||| | ||| |  | || |||||     |||    |||| |  || ||    || |
Db      29738 CTCGCCGAGCAGCCCGCCGCGGACCACGACGCCATCCTGCTGGCACTGGTCACCGAGAAG 29797

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |  ||| | || ||| | ||||| ||||    |      |||  ||||||| | | |||||
Db      29798 GCCGCGACCGTCCTCGGCCACGCCGACCACGACCTGCTCGAACCCGACCTGCCCTTCCGC 29857

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |  || |||||||| || || || || |||||  | || |||| ||| | |  || |||
Db      29858 GACCTGGGCTTCGACTCGCTGACCGCGGTCGACCTGCGCAACCAGCTCACCGCGGAAACC 29917

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || |||  ||| || | | || ||||||||||||| | ||| |||   ||||||
Db      29918 GGCCTCACCCTGCCCGCCACCCTCGTCTTCGACCACCCCAACCCGGCCGCCCTCGCCGCC 29977

Qy        241 CACCTG 246
              ||||||
Db      29978 CACCTG 29983


RESULT 8
AAA92302
ID   AAA92302 standard; DNA; 31422 BP.
XX
AC   AAA92302;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag= a
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             14824..31422
FT                   /*tag= b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23751, AAB23752.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use -
XX
PS   Claim 2; Page 134-203; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;

  Query Match             43.7%;  Score 107.4;  DB 21;  Length 31422;
  Best Local Similarity   64.9%;  Pred. No. 2.3e-19;
  Matches  159;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              ||||||   |||||    |   | || |   || |  | ||||  || || ||   |||
Db      20089 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 20148

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              | ||  | || ||| | ||| |   |||  |  ||||| | |||||||   |||||||
Db      20149 ATCGCCACCGTCCTCGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 20208

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |  |||||||||| || ||||| ||||||||  | || |||| |||   |    | ||
Db      20209 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 20268

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              |||||||||| ||   | | || | |||||||||||| | |||  |||  ||| ||
Db      20269 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 20328

Qy        241 CACCT 245
              |||||
Db      20329 CACCT 20333


RESULT 9
AAH79278
ID   AAH79278 standard; DNA; 31422 BP.
XX
AC   AAH79278;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 2.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag= a
FT                   /product= "AAG65266"
FT   CDS             14824..31422
FT                   /*tag= b
FT                   /product= "AAG65267"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
```

```
PR   29-OCT-1998;   98US-0106100.
PR   16-FEB-1999;   99US-0120254.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Betlach MC,  Shah SK,  McDaniel R,  Tang L;
XX
DR   WPI; 2000-365602/31.
DR   P-PSDB; AAY92707, AAY92708, AAY92709.
XX
PT   Recombinant DNA compound encoding oleandolide polyketide synthase for
PT   synthesizing polyketides comprising a coding sequence for a domain of a
PT   loading module or any one of extender modules
XX
PS   Disclosure; Page 14-26; 86pp; English.
XX
CC   This is part of the Streptococcus antibioticus oleandomycin gene cluster.
CC   The oleandolide polyketide synthase (PKS), also known as
CC   8,8a-deoxyoleandolide synthase, is encoded by three open reading frames
CC   (ORF), designated oleAI, oleAII and oleAIII. The PKS is a type I
CC   "modular" enzyme, where each ORF encodes 2 extender modules and
CC   the first ORF also encodes the loading module. Each module is composed
CC   of at least a ketosynthase (KS), acyl-transferase (AT) and an
CC   acyl carrier protein (ACP) domain. The oleandolide PKS loading module
CC   contains an inactivated KS, called KS-Q, where Q is the abbreviation for
CC   glutamine, present instead of the active site cysteine required for
CC   activity. The large multifunctional PKS enzymes catalyze the biosynthesis
CC   of polyketide macrolactones through multistep pathways involving
CC   decarboxylative condensations between acylthioesters followed by cycles
CC   of varying beta-carbon processing activities. The macrolide product of
CC   the PKS, 8,8a-deoxyoleandolide, is further modified by epoxidation and
CC   glycosylation to  yield oleandomycin, an antibacterial polyketide. The
CC   invention concerns an isolated recombinant DNA compound, comprising a
CC   coding sequence for a domain of loading module or any one of extender
CC   modules 1-4 or 1-6, including an oleandolide PKS operably linked to a
CC   promoter. Also discussed are recombinant oleandolide PKS in which the
CC   module 1 KS domain is inactivated by deletion or other mutation. In
CC   particular, the inactivation is mediated by a change in the KS domain
CC   that renders it incapable of binding substrate (the KS1-o mutation),
CC   rendered by mutation in the codon for the active site cysteine. The
CC   oleandolide PKS is useful for synthesizing polyketides, which are useful
CC   as antibiotics and motilides. Heterologous expression of oleandolide PKS
CC   in host cells such as Streptomyces coelicolor and S. lividans is also
CC   made possible. Unmodified oleandolide compounds can be provided to
CC   cultures of Saccharopolyspora erythraea and converted to the
CC   corresponding derivatives of erythromycins A-D.
XX
SQ   Sequence 50937 BP; 6672 A; 16253 C; 19272 G; 8740 T; 0 other;

  Query Match             44.8%;  Score 110.2;  DB 21;  Length 50937;
  Best Local Similarity   66.8%;  Pred. No. 4.5e-20;
  Matches  157;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy         12 GACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCATGTGGCGGCAGT 71
              | || |  | |||| |||    ||  ||| ||| ||||| || | ||| || ||    ||
Db      29150 GGCGGCACCGGACCCACGTCGTGCCCTGCGCGACCTCGTCCGCGGCCACGTCGCCCTGGT 29209

Qy         72 GCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGTGCACTCGGCTT 131
              |||| | |||   | |||  | |||||||    |||| ||| | ||||| |   |||| ||
Db      29210 GCTCGGACACAGTGGCCCCGAGGCCATCGACGCCGAACAGGCCTTCCGGGACATCGGTTT 29269

Qy        132 CGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACAGGACTCCGCCT 191
              ||| || || || || ||||| || |||||| ||||| |   |  || || ||||| |
Db      29270 CGACTCCCTGACCGCAGTCGAACTCAGAAACCGGCTGAACGCCGAGACCGGCCTCCGCTT 29329

Qy        192 TCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTACACCTG 246
               || | | |||| || |||||| |||| | |||   |   ||||||| ||||||
Db      29330 GCCCGGCACGCTCGTGTTCGACTACCCCAACCCGAGCGCGCTCGCCGATCACCTG 29384


RESULT 7
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag=  a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag=  b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag=  c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag=  d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag=  e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag=  f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag=  g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag=  h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
```

```
RESULT 5
AAF24892
ID   AAF24892 standard; DNA; 20394 BP.
XX
AC   AAF24892;
XX
DT   20-APR-2001  (first entry)
XX
DE   Pimaricin biosynthesis associated polyketide synthase gene.
XX
KW   Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW   anticancer; pimaricin; ss.
XX
OS   Streptomyces natalensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..20394
FT                   /*tag= a
FT                   /product= "polyketide synthase"
XX
PN   WO200077222-A1.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-EP06227.
XX
PR   14-JUN-1999;   99EP-0201893.
XX
PA   (STAM ) DSM NV.
XX
PI   Martin JF,  Aparicio JF,  Colina AJ;
XX
DR   WPI; 2001-080693/09.
DR   P-PSDB; AAB31558.
XX
PT   New polynucleotides encoding enzymes involved in the biosynthesis of
PT   pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT   the synthesis of new compounds  -
XX
PS   Disclosure; Page 53-80; 116pp; English.
XX
CC   The present sequence encodes a polyketide synthase which is associated
CC   with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC   is useful for the oxidative modification of a methyl group of a suitable
CC   compound, e.g. a bioactive compound including a secondary metabolite,
CC   antibiotics and anticancer agents. Recombinant cells comprising the
CC   gene are useful for the production of pimaricin. The polyketide synthase
CC   polynucleotide may be over expressed in Streptomyces, leading to an
CC   increase in the biosynthesis of pimaricin, as a source of primers for
CC   amplification reaction and as probes.
XX
SQ   Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;

  Query Match           45.6%;  Score 112.2;  DB 22;  Length 20394;
  Best Local Similarity   66.1%;  Pred. No. 1.1e-20;
  Matches  162;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||   | | |  |||| || |  || |  ||  |||||||  ||||||| | |||
Db      14677 CTTGCCGGGCTGTCCACGGCGGAGCGGCGCGAGGCGCTGCTCGATGTCGTACGGGCCCAG 14736

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
               | ||| | || ||  | ||||| | |||| || | ||||| || |||| | | ||||
Db      14737 ATCGCGACGGTCCTGGGCCACGCCGGCCCGGAAACGATCGCCCCTGACCGGGCCTTCCAG 14796

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |  |||||| |||| || || |||||  ||||  |||| ||||||||   |||||| ||
Db      14797 GACCTCGGCCTCGACTCCCTGACGGCGATCGAACTCCGTAACCTGCTCGGCAAGGCCACC 14856

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || |   ||   || |||||| |||||||     |     || ||||
Db      14857 GGGCTGCGGCTCCCGGCAACGACCGTGTTCGACTACCCGACGGTGGATGCCCTGGCCGCC 14916

Qy         241 CACCT 245
               |||||
Db      14917 CACCT 14921


RESULT 6
AAA09469
ID   AAA09469 standard; DNA; 50937 BP.
XX
AC   AAA09469;
XX
DT   29-AUG-2000  (first entry)
XX
DE   Streptococcus oleandomycin gene cluster.
XX
KW   oleandomycin; oleandolide; polyketide synthase; oleAI; oleAII; oleAIII;
KW   PKS; type I; 8,8a-deoxyoleandolide synthase; modular; ketosynthase;
KW   acyl-transferase; acyl carrier protein; inactivated; polyketide;
KW   macrolactone; antibiotic; motilide; erythromycin; ss.
XX
OS   Streptococcus antibioticus.
XX
FH   Key             Location/Qualifiers
FT   CDS             152..1426
FT                   /*tag= a
FT                   /label= oleI
FT   CDS             complement (1528..2637)
FT                   /*tag= b
FT                   /label= oleN2
FT   CDS             complement (2658..4967)
FT                   /*tag= c
FT                   /label= oleR
FT   CDS             5772..18224
FT                   /*tag= d
FT                   /label= ORF1
FT                   /product= 8,8a-deoxyoleandolide_synthase_1
FT   CDS             18267..29717
FT                   /*tag= e
FT                   /label= ORF2
FT                   /product= 8,8a-deoxyoleandolide_synthase_2
FT   CDS             29787..40346
FT                   /*tag= f
FT                   /label= ORF3
FT                   /product= 8,8a-deoxyoleandolide_synthase_3
FT   CDS             40625..41830
FT                   /*tag= g
FT                   /label= oleP1
FT   CDS             41878..43158
FT                   /*tag= h
FT                   /label= oleG1
FT   CDS             43163..44443
FT                   /*tag= i
FT                   /label= oleG2
FT   CDS             44433..45173
FT                   /*tag= j
FT                   /label= oleM1
FT   CDS             45251..46411
FT                   /*tag= k
FT                   /label= oleY
FT   CDS             46491..47714
FT                   /*tag= l
FT                   /label= oleP
FT   CDS             complement (47808..49517)
FT                   /*tag= m
FT                   /label= oleB
XX
PN   WO200026349-A2.
XX
PD   11-MAY-2000.
XX
PF   22-OCT-1999;   99WO-US24478.
XX
```

```
FT   CDS             11971..30690
FT                   /*tag=  b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23749, AAB23750.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 2; Page 66-134; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;

  Query Match             100.0%;  Score 246;  DB 21;  Length 30690;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-56;
  Matches  246;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5935 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 5994

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5995 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 6054

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6055 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 6114

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6115 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 6174

Qy        241 CACCTG 246
              ||||||
Db       6175 CACCTG 6180


RESULT 4
AAH79277
ID   AAH79277 standard; DNA; 30690 BP.
XX
AC   AAH79277;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 1.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag=  a
FT                   /product= "AAG65264"
FT                   /partial
FT   CDS             11971..30690
FT                   /*tag=  b
FT                   /product= "AAG65265"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65264, AAG65265.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Example 2; Page 58-123; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;

  Query Match             100.0%;  Score 246;  DB 22;  Length 30690;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-56;
  Matches  246;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5935 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 5994

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5995 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 6054

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6055 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 6114

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6115 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 6174

Qy        241 CACCTG 246
              ||||||
Db       6175 CACCTG 6180
```

```
PI   Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65268.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 10; Page 149-167; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is modified version of a
CC   fragment of the S. avermitilis genome.
XX
SQ   Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;

  Query Match             100.0%;  Score 246;  DB 22;  Length 11916;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-56;
  Matches  246;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5935 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 5994

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5995 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 6054

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6055 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 6114

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6115 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 6174

Qy        241 CACCTG 246
              ||||||
Db       6175 CACCTG 6180


RESULT 2
AAZ58381
ID   AAZ58381 standard; DNA; 12381 BP.
XX
AC   AAZ58381;
XX
DT   23-MAY-2000  (first entry)
XX
DE   Streptomyces avermitilis avermectin polyketide synthase modules 1+2.
XX
KW   Polyketide synthase; avermectin; insecticide; ss.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200001827-A2.
XX
PD   13-JAN-2000.
XX
PF   06-JUL-1999;   99WO-GB02158.
XX
PR   06-JUL-1998;   98GB-0014622.
XX
PA   (BIOT-) BIOTICA TECHNOLOGY LTD.
PA   (PFIZ ) PFIZER INC.
XX
PI   Kellenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ;
PI   McArthur HAI;
XX
DR   WPI; 2000-182117/16.
XX
PT   Mutated Type I polyketide synthase containing a polylinker site in an
PT   extension module for replacement of a reductive loop sequence, for
PT   producing polyketides, e.g. B1 avermectin -
XX
PS   Disclosure; Fig 7a-f; 75pp; English.
XX
CC   The present sequence is that of DNA encoding the first 2 modules
CC   of the avermectin polyketide synthase (PKS) of Streptomyces
CC   avermitilis.  The invention relates to nucleic acids encoding a
CC   Type I PKS such as avermectin in which a polylinker with multiple
CC   restriction sites replaces or 1 more PKS genes encoding enzymes
CC   associated with reduction.  Novel PKS are provided in which in
CC   which the reductive loop in a selected module of the Type I PKS is
CC   replaced with the equivalent segment from the same or different
CC   PKS gene cluster or by a mutated or synthetic segment.  Vectors and
CC   host cells, and methods for producing novel polyketides by
CC   culturing host cells are claimed.  The polyketides obtained are
CC   useful as antibiotics and insecticides.  Fermentation products
CC   containing C22-C23 dihydroavermectin, ivermectin and B1
CC   avermectins are claimed.
XX
SQ   Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;

  Query Match             100.0%;  Score 246;  DB 21;  Length 12381;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-56;
  Matches  246;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6752 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 6811

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6812 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 6871

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6872 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 6931

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6932 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 6991

Qy        241 CACCTG 246
              ||||||
Db       6992 CACCTG 6997


RESULT 3
AAA92301
ID   AAA92301 standard; DNA; 30690 BP.
XX
AC   AAA92301;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag=  a
FT                   /note= "avermectin aglycon synthase protein"
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 48.936 Seconds
(without alignments)
11320.736 Million cell updates/sec

Title: US-09-914-286-1_COPY_5935_6180
Perfect score: 246
Sequence: 1 ctggcccatcagacgtcggc.....ccaaactcgccgtacacctg 246

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	246	100.0	11916	22	AAH79279	Streptomyces averm
2	246	100.0	12381	21	AAZ58381	Streptomyces averm
3	246	100.0	30690	21	AAA92301	S. avermitilis ave
4	246	100.0	30690	22	AAH79277	Streptomyces averm
5	112.2	45.6	20394	22	AAF24892	Pimaricin biosynth
6	110.2	44.8	50937	21	AAA09469	Streptococcus olea
7	110	44.7	125401	22	AAD17186	Streptomyces nours
8	107.4	43.7	31422	21	AAA92302	S. avermitilis ave
9	107.4	43.7	31422	22	AAH79278	Streptomyces averm
10	105.2	42.8	65140	22	AAD17184	Streptomyces nours
11	103.6	42.1	1681	24	AAS18438	Contig 115 DNA enc
12	103.6	42.1	11220	21	AAZ87298	S. venezuelae macr
13	103.6	42.1	36778	21	AAZ87318	S. venezuelae pik
14	103.6	42.1	37948	21	AAZ87285	S. venezuelae pik
15	103.6	42.1	38506	21	AAA75633	Nucleotide sequenc
16	103.6	42.1	38506	21	AAZ56001	Recombinant cosmid
17	102	41.5	43280	18	AAT80413	Tylactone synthase
18	102	41.5	53789	19	AAV21187	Amycolatopsis medi
19	101	41.1	77536	21	AAA14651	Nucleotide sequenc
20	100.6	40.9	44377	18	AAT78508	Platenolide syntha
21	100.6	40.9	44377	18	AAT80414	Platenolide syntha
22	100.4	40.8	2700	22	AAD17193	ERD48 insert DNA i
23	100.4	40.8	9513	22	AAF88337	S. spinosa DNA fra
24	100.4	40.8	50000	22	AAF88313	S. spinosa DNA fra
25	100.4	40.8	50000	22	AAF88316	S. spinosa DNA fra
26	100.4	40.8	80161	20	AAZ21501	DNA fragment of Sa
27	98.8	40.2	5676	19	AAV21186	Amycolatopsis medi
28	98.4	40.0	4689	21	AAZ87299	S. venezuelae macr
c 29	97.8	39.8	1434	24	AAS18434	Contig 132 DNA enc
c 30	97.6	39.7	77536	21	AAA14651	Nucleotide sequenc
31	96.8	39.3	6459	22	AAF88336	S. spinosa DNA fra
32	96.2	39.1	15872	18	AAT68715	Streptomyces venez
33	96.2	39.1	15872	21	AAZ87283	S. venezuelae vep
34	95.6	38.9	27541	22	AAD17185	Streptomyces nours
35	94.6	38.5	13842	21	AAZ87297	S. venezuelae macr
36	94.6	38.5	29879	14	AAQ46806	eryA region of S.
c 37	92.8	37.7	561	22	AAA89135	Polyketide synthas
38	92.4	37.6	16767	22	AAF88339	S. spinosa DNA fra
39	92.4	37.6	25360	22	AAF88314	S. spinosa DNA fra
40	92.4	37.6	29736	22	AAF88317	S. spinosa DNA fra
41	89	36.2	13987	18	AAT80415	Hybrid srmG/tylG O
42	88	35.8	47981	22	AAF30757	Micromonospora meg
43	87.8	35.7	14775	22	AAF88338	S. spinosa DNA fra
44	86	35.0	4209	22	AAH51966	Mycobacterium tube
45	86	35.0	4209	22	AAF31625	Mycobacterium tube

ALIGNMENTS

RESULT 1
AAH79279
ID AAH79279 standard; DNA; 11916 BP.
XX
AC AAH79279;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis coding sequence derivative SEQ ID NO: 3.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..11916
FT /*tag= a
FT /product= "AAG65268"
FT /partial
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX

THIS PAGE BLANK (USPTO)

```
                AGAEPLPLDGHYDRLADGGFGYGPVFQGLRAAWRGGDVVYAEVELPEAGRSDAEAFGL
                HPALLDAALHAAPFTGLGERGRGGLPFSWEGVSLHAGGATTLRVRLTPVADDALALTV
                ADGTGAPVLSVDSLVLRSVATQQLDTAAAVARDALFRLDWTPVQPTATDPGPVALLGA
                DPFGLLTHAGFADAPAYPDLAALAAADGPVPTTVVLSLAGTGDDAADPARSAHRCAAE
                ALAAVQTWLDHHERFAAARLVFVTRGATVGRDVAAAAVWGLVRSAQSENPGCFALVDL
                DPDGAVGAAALVAALVSGEPQLAVRGDVLRVARLVRRPLTEVGAGADGTGDGVGDGSG
                VSFSGEGAVLVTGGTGGLGAVLARHLVAEYGVRDLLLVSRSGERAVGAGELVAELAGV
                GARVRVVACDVTDRAAVVELVGGHAVSAVVHAAGVLDDGMVGALTGERLSAVLRPKVD
                AVWHLHEATRGLDLDAFVVFSSLAGVFGSPGQANYAAANAFLDALMTRRRAEGLPGLS
                LAWGPWEQSGGMTGTLTDVDAERLARSGVPPLSVAQGLALFDAAVAGTDATCVPVRLD
                LPVLRARGEVPPLLRSLIRVRARRAAVAGSATAGNLAQRLRRLDEDGRDEMVLDLVRG
                QVALVLGHATGGDVDAGRAFRDLGFDSLTAVELRNRLNTVTGLRLPATLVFDYPTVRH
                LATYVLDELLGTDAEVATVQPAAVAVADDPIVIVGMACRYPGGVSSPEDLWRVLTEGT
                DAVSGFPTNRGWDVESLYHPDPDHPGTSYTRSGGFLHEAGEFDPGFFGMSPREALATD
                SQQRLLLESSWEAIERAGIDPVSLRGSRTGVFAGVMYSDYSAMLASPEFEGFQGSGSS
                PSLASGRVAYTLGLEGPAVTVDTACSSSLVAMHWAMQALRSGECGLALAGGVTVMSTP
                AVFVDFARQRGLSPDGRCKAFADAADGVGWSEGVGVLVLERQSDAVRNGHEILAVVRG
                SAVNQDGASNGLTAPNGPSQQRVIRQALASGGLTAGDVDVVEAHGTGTTLGDPIEAQA
                LLATYGRDREPERPLLLGSVKSNLGHTQAAAGVAGVIKMVLAMRHGVVPRTLHVDAPS
                SHVDWSEGAVELLSEQAAWPETGRVRRAGVSSFGISGTNVHVIVEQAPGAKAIAAAGA
                ARRTPGAVPVLLSGRGRSALRGQAARLLGHLQARPDAELVDVALSLATTRSRFEQRAA
                VVAQDRDQLIASLGALAADRPDPAVVEGEAAGRGRTAVLFTGQGSQRAAMGRELHEVQ
                PEFAAAFDAVCAVFDPLLDRPLREVVFAEDGSDEAALLDETGWTQPALFAVEVALFRL
                VESWGVRPDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARATLMQALPTGGAMIAIQA
                TEDEIAAHLDDTVAIAAVNGPQSVVISGDEEAAETIAATFAERGRKTKRLRVSHAFHS
                PRMDGMLDAFRIVAEGLTYRAPRIPLVSDLTGRRADDAEVCTAEYWVRHVREAVRFAD
                CVRTLRDAGATTFLELGSDGLLTAMAEDTLGDDHDAELVPMLRAGRAEELAAATALAR
                LQVRGVDVDWAAYLAGTGARRTDLPTYAFQHAYYWPQLPTPAAALAAADPADQQLWAA
                VERGDARELADILGLGEQDLTPLDSLLPALTSWRRGNQEKHLLDTLRYRVEWTRLSKP
                TAPVLDGTWLLVASDATAADQPALLDGLADALGSHGARVRRLLLDDSCADRAVLAERL
                ARTADVDAATQVLSVLPLDERDADDCPPLTRGLALTVALVQALADTGAQGRLWTATRG
                AVSTNPADPVTHPVQAAAWGLGRGVALEHPRLWGGLVDLPQVFDERAGQRLAGILAVK
                DAPDGEDQVALRATGVSGRRLVRHTVEALPTAAEFTATGTVLITGGTGGLGAEVARWL
                ARAGAQHLVLTSRRGPDAPGAAELRAELEGYGPSVSVVACDVADRDALAAVLTALPEE
                LPLTGVVHTAGVGHYGPLDTLSTAEFAGLTAAKLAGAAHLDALLADRELDFFVLFGSI
                AGVWGSGNQSAYGAANAYLDALALHRRARGLAATSVAWGPWAEAGMAADDAVSETLRR
                QGLGLLDPAPAMTELRRAVVRQDVTVTVADVDWQRYAPLFTSARPSALIAGLPEVRAL
                AADERTEQDATGASEVVTRVRALAEPEQLRLLTDLVRTESATVLGHSSADAVPEGRAF
                RDVGFDSLTAVELRKRLGAATGLSLPSTMVFDYPTPLELAQYLRAEILGAVLEVAGPV
                ATGGADDEPIAIIGMACRFPGGVSSPEQLWDLVASGTDAISEFPVNRGWQTGHLFDPD
                PDRPGTTYSTQGGFLHEADEFDPTFFGISPREALVMDPQQRLLLETTWESFERAGIRP
                ETLRSTLTGTFVGSSYQEYGLGAGDGTEGHMVTGSSPSVLSGRLSYVFGLEGPAVTVD
                TACSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNPFIAFSRQRALAKDGRCKAFS
                DDADGMTLAEGVGVVLVERLSDAQRNGHPVLAVLRGSAINQDGASNGLTAPNGPSQQR
                VIRQALANARLAPGDIDALEAHGTGTPLGDPIEAQALFATYGRDRDPESALLLGSVKS
                NIGHTQSAAGIASVIKMVMALRHSELPPTLHADAPSSHVDWSAGTVRLLTQARAWPET
                GRPRRAAVSSFGISGTNAHVLLEQAPVADTPAEERPAVAPVPIAAGVVPWVVTARSAA
                ALRGQAERLLAHAETVGTALPAAGPLDIGLSLVSARARFEHRAVVVPPAGTDPLAALR
                AVATDGPSPVVARGVADVEGRTVFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAAL
                AEFTDWSLVDVLRGVVGAPSLERVDVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQG
                EIAAAVVSGALSLRDGARVVALRSQAIGRALAGRGGMMSVALSVDVLEPRLVEFEGRV
                SVAAVNGPRSVVVAGEPEALDALHARLTADDIRARRIAVDYASHSHQVEDLHEELLEV
                LAELAPRTSEVPFFSTVTGDWLDTARMDAGYWFRNLRGRVRFADAVADLLAAEYRAFV
                EVSSHPVLSMAVQEAIDEAGVPAVAAGTLRRDQGGTDRFLLSAAEVFVRGVDVDWAGL
                FEGTGASRIDLPTYAFQHEHLWAVPPAPEAVAAADPDDAAFWTAVEDGDVSALTAALG
                TDEDSVAAVLPALTSWRRARRDRSTVDAWRYRVAWKPLGGTLPHPSLTGTWLLVTADG
                IDDTDVAGALETYGAEVRRLVLDEECVDRAVLRERLAGAEDVTGIVSVLAAAERTDAV
                PGTSLVLGTALTVALIQALGDAEIDAPVWALTRGAVSTGRADELTAPVQAQVTGIGWT
                AALEHPQRWGGTLDLPAALDARAAQRLAAVLSGALGSDDQLAIRPSGVFTRRIVRAEA
```

Query Match 44.7%; Score 110; DB 1; Length 123580;
Best Local Similarity 65.4%; Pred. No. 8.1e-14;
Matches 161; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

```
Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || ||| | ||| |  | || |||||     |||    |||| |  || ||    || |
Db      29330 CTCGCCGAGCAGCCCGCCGCGGACCACGACGCCATCCTGCTGGCACTGGTCACCGAGAAG 29389

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |  ||| | || ||| | ||||| ||||   |     |||  ||||||| | | |||||
Db      29390 GCCGCGACCGTCCTCGGCCACGCCGACCACGACCTGCTCGAACCCGACCTGCCCTTCCGC 29449

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |  || |||||||| || || || || |||||  | || |||| ||| | |  || |||
Db      29450 GACCTGGGCTTCGACTCGCTGACCGCGGTCGACCTGCGCAACCAGCTCACCGCGGAAACC 29509

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || |||  ||| || | | | || |||||||||||||| | ||| |||   |||||||
Db      29510 GGCCTCACCCTGCCCGCCACCCTCGTCTTCGACCACCCCAACCCGGCCGCCCTCGCCGCC 29569

Qy        241 CACCTG 246
              ||||||
Db      29570 CACCTG 29575
```

Search completed: June 17, 2003, 17:55:36
Job time : 479.645 secs

```
Db      29330 GCCCGGCACGCTCGTGTTCGACTACCCCAACCCGAGCGCGCTCGCCGATCACCTG 29384


RESULT 15
AF263912
LOCUS       AF263912               123580 bp    DNA     linear . BCT 24-MAY-2000
DEFINITION  Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF263912
VERSION     AF263912.1  GI:8050835
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., StrLm,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Biosynthesis of the polyene antifungal antibiotic nystatin in
            Streptomyces noursei ATCC 11455: analysis of the gene cluster and
            deduction of the biosynthetic pathway
  JOURNAL   Chem. Biol. 7 (6), 395-403 (2000)
  MEDLINE   20334850
   PUBMED   10873841
REFERENCE   2  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
            N-7489, Norway
FEATURES             Location/Qualifiers
     source          1. .123580
                     /organism="Streptomyces noursei"
                     /strain="ATCC 11455"
                     /db_xref="ATCC:11455"
                     /db_xref="taxon:1971"
     gene            complement(46. .783)
                     /gene="nysF"
     CDS             complement(46. .783)
                     /gene="nysF"
                     /function="presumably post-translationally modifies ACP
                     domains on PKS"
                     /note="putative 4'-phosphopantheteine transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysF"
                     /protein_id="AAF71762.1"
                     /db_xref="GI:8050836"
                     /translation="MIELILPATVATEAAYDDRPRPGDRLLSSEREVIARAVESRQRE
                     FTTVRHLARRALRRLGHPDRAILPNRRGAPQWPPGIVGSMTHCAGYRAAAVSPAELSA
                     AVSIDAEPNGPLPAGVLNAIALPSERPHLVALAAHRPDVHWDRLLFSAKESVFKAWYP
                     LTQRELDFSEAEIVIDPTQGAFTARLLVPGPLLGGRRVTVFPGRWHSTPALLTTAVHL
                     PAPTPRRDREHRTHLTVNSPLPRPTFG"
     gene            complement(867. .2684)
                     /gene="nysG"
     CDS             complement(867. .2684)
                     /gene="nysG"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysG"
                     /protein_id="AAF71763.1"
                     /db_xref="GI:8050837"
                     /translation="MASPDDLEEERTAPRPARRLVGLLRPHRRSVALAVSMGVGGIVL
                     NAFGPLLLGRVTDLIADGVLGGVPGPAPGIDFAAIGRLLLVLLALYVVASLFMLAQGR
                     LVASAVWRTIHELRRDAREKLTRLPLRHFDRQPAGELLSRTTNDIDNLQQTLQQTLAE
                     LITSIFSLLTMLVLMLVISPSLAVVMLLSVPVSALIAARISKRAQPHYAAQWSANGTL
                     NAHVEEVCTGHALIKGFDRRAAAEERFDACNDAVYRAAAKAQFASGAMEPVMMFVANL
                     GYVLVAVIGAWKVINGTLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGIASAQRVF
                     TLLDAPEQAPDPLRPGTPARAEGRVEFTDVSFRYSPDTPLIENLSLTVEPGSTVAIVG
                     PTGAGKTTLGNLLMRFYEPDSGRILLDGTDTATMTRDDLRSRFGLVLQDTWLFGGTIA
                     ENIAYGAPGACRADIEEAARATCADRFIRTLPQGYDTVLDDESGTVSAGEKQLLTVAR
                     AFLARPAVLVLDEATSSVDTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIVVM
                     DAGRIVEQGTHDQLLCAQGLYARLHAARTHTPTAGAAAG"
     gene            complement(2662. .4416)
                     /gene="nysH"
     CDS             complement(2662. .4416)
                     /gene="nysH"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysH"
                     /protein_id="AAF71764.1"
                     /db_xref="GI:8050838"
                     /translation="MLLRLLRAQLRPYAWATAALVALQLVQILGTLLLPTLGAALIDQ
                     GVVRGDGGRITELGVVMGVVALVQIAAALGAAALAARTATAMGRDLRSALFRRILDFS
                     AREIGRFGTPSLLTRSVNDVQQVQNLAQTGFGIVVCAPLMCLGSVLLALRQDVPLALL
                     LVALVLVVAVCFGLLLARMGTLYARMQLTLDRLGRLLREAITGVRVVVRSFVRDDHERA
                     RFAQTNDAFLVVSRRVGRLIATMLPVVLLLMNGFTVALLWTGSHRIDAGRMPIGSLSA
                     LLSYLSLILMSVVMLAFVFLSVPRARVCAGRIAEVLDTGSSVAPPAAPQPVRGPAGRI
                     ELCAAGYRYPGAEEPVLRDVDLTVEPGERIAVLGSTGSGKTTLLNLVLRLADATEGAV
                     RVGGTDVRELTAATLAAAVGFVPQRPYLFSGTVASNLRFGRPDATDEELWEALRVAQA
                     ADFVARMPDGLDAEITQGGGNVSGGQRQRLSLARALLRRPEIYLFDDCFSALDQATDA
                     ALRTALVPYTAGATVITVAQRISAGRDADRIVVLDRGRVVAQGTHDVLLRTSPTYREI
                     ALSQLTEEEAAHGLAGRP"
     gene            4714. .5748
                     /gene="nysDIII"
     CDS             4714. .5748
                     /gene="nysDIII"
                     /function="presumably involved in mycosamine biosynthesis"
                     /note="putative dGDP-mannose-4,6-dehydratase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysDIII"
                     /protein_id="AAF71765.1"
                     /db_xref="GI:8050839"
                     /translation="MSKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKSRV
                     SRLASELDFIDGDLMDQGSLVSAVDTVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                     VLRMLEAIRMVSGLSTSRTVSPRGQIRFYQASSSEMFGKAAETPQRETTLFHPRSPYG
                     AAKAYGHYITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKQGLQDKL
                     ALGNLDAVRDWGYAGDYVRAMHLMLQQDAGDDYVIGTGQMHSVRDAVRIAFEHVGLNW
                     EDYVVIDPDLVRPAEVEVLCADSAKAQDRLGWKPDVDFPTLMRMMVDSDLAQVSRENQ
                     YGDVLLAANW"
     gene            5930. .34363
                     /gene="nysI"
     CDS             5930. .34363
                     /gene="nysI"
                     /function="responsible for condensation steps 9 to 14 in
                     the nystatin polyketide backbone synthesis"
                     /note="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysI"
                     /protein_id="AAF71766.1"
                     /db_xref="GI:8050840"
                     /translation="MDNEQKLRDYLKLATADLRRTRRRVHKLESAAQEPVAIIGMTCR
                     YPGGVRSPEDLWRMVEAGEHGVTPFPTDRGWDLEALAAAPTASGGFLHDAPDFDADFF
                     GISPREAVAMDPQQRVVLESAWEAFERAGIDPTSVKGSRTGVFIGAMAQDYRVGPADG
                     AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                     AGGVTIMSGPGTFIEMGRQGGLSADGRCRSFGDTADGTGWAEGVGILVLERLSDAVRN
                     GHEILAVVRGTAVNQDGASNGLTAPNGPSQQQVIQQALVNARLAAGDIDVVEAHGTGT
                     TLGDPVEAQALLATYGQNRPADRPLLLGSVKSNLSHTQAAAGVAGVIKMVMAMRHGTL
                     PRTLHAEEPTHHVDWSQGAVRLLTDTTDWPATGAPRRAAVSSFGISGTNAHTIIEQAP
                     EPQPEDAATAQDDAAGSTPATAPVVPGVVPVLLSGRTPDALRGQAAALRAALDTGRRP
                     DLLDLAHSLATTRAGFEHRAVLLATDHPALTDGLTALADADDPAAAPAWITGTTRAET
                     RLAVLFTGQGAQRLGAGRELAARFPAFATALDAALDAFTPHLDRPLREVLWGTDAALL
                     DRTAYAQPALFAVEVALYRLIESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAATLVA
                     ARGRLMQALPDGGAMIAVQASEADVAPLLAGHEDQVAIAAVNGPSAVVLSGAEATVTA
                     LAEQLAADGRKTRRLRVSHAFHSPLMEPMLDAFRAVVEDLTLQPPLLPVVSNLTGKPA
                     TVAQLTSADYWVDHVRHAVRFADGIDWLARHDTTAFLELGPDGVLSAMAQDCLDAADA
                     DAVTLPALRAGRPEEHTLTTALAGLHVHGATLDWTGCFAGTGARRTDLPTYAFQRRRY
                     WPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRG
                     TTLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRT
                     VNVHARPDTGDDTPWTAHATGVLTTADASRQLPASSEQGGTPLAGDPHPALDAAQWPP
```

```
                     /note="Region: acyltransferase domain AT-L"
misc_feature         3485. .3724
                     /gene="oleAI"
                     /note="Region: acyl carrier protein domain ACP-L"
misc_feature         3806. .5056
                     /gene="oleAI"
                     /note="Region: ketoacyl synthase domain KS1"
misc_feature         5363. .6400
                     /gene="oleAI"
                     /note="Region: acyltransferase domain AT1"
misc_feature         7109. .7669
                     /gene="oleAI"
                     /note="Region: ketoacyl reductase domain KR1"
misc_feature         7943. .8200
                     /gene="oleAI"
                     /note="Region: acyl carrier protein domain ACP1"
misc_feature         8258. .9541
                     /gene="oleAI"
                     /note="Region: ketoacyl synthase domain KS2"
misc_feature         9848. .10882
                     /gene="oleAI"
                     /note="Region: acyltransferase domain AT2"
misc_feature         11723. .12274
                     /gene="oleAI"
                     /note="Region: ketoacyl reductase domain KR2"
misc_feature         12560. .12817
                     /gene="oleAI"
                     /note="Region: acyl carrier protein domain ACP2"
gene                 13118. .24568
                     /gene="oleAII"
CDS                  13118. .24568
                     /gene="oleAII"
                     /note="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="8,8a-deoxyoleandolide synthase 2"
                     /protein_id="AAF82409.1"
                     /db_xref="GI:9049536"
                     /translation="MTNDEKIVEYLKRATVDLRKARHRIWELEDEPIAITSMACHFPG
GIESPEQLWELLSAGGEVLSEFPDDRGWDLDEIYHPDPEHSGTSYVRHGGFLDHATQF
DTDFFGISPREALAMDPQQRLLLETSWQLFERAGVDPHTLKGSRTGVFVGAAHMGYAD
RVDTPPAEAEGYLLTGNASAVVSGRISYTFGLEGPAVTVDTACSSSLVALHLAVQALR
RGECSLAVVGGVAVMSDPKVFVEFSRQRGLARDGRSKAFAASADGFGFAEGVSLLLLE
RLSDARRLGHRVLAVVRGSAVNQDGASNGLAAPNGPSQQRVIRAALADAGLAPADVDV
VEAHGTGTRLGDPIEAQALLATYGQGRTSGRPVWLGSVKSNIGHTQAAAGVAGVMKMV
LALERGVVPKTLHVDEPSPHVDWSTGAVELLTEERPWEPEAERLRRAGISAFGVSGTN
AHVIVEEAPAEPEPEPEPGTRVVAAGDLVVPWVVSGRDAGALRAQAARLAAHVSSTGA
GVVDVGWSLVATRSVFEHRAVMVGTDLDSMAGSLAGFAAGGVVPGVVSGVAPAEGRRV
VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPLTGWSLVEVLRGGEAVLGRV
DVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRS
RAIARIAGGGGMVSVSLPAGRVRTMLDTYGGRLSVAAVNGPSSTVVSGDAQALDELLA
GCEREGVRARRVPVDYASHSAQMDQLRDELLEALADITPQHSSVPFFSTVTADWLDTT
ALDAGYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDTVEADAVA
LGSLRRDEGGLGRFLTSLAEAFVQGVPVDWSRTFEGASPRTVDLPTYPFQRQRFWLEG
SPALSSNGVEGEADVAFWDAVEREDSAVVAEELGIDAKALHMTLPALSSWRRRERQRR
KVQRWRYRVEWKRLPNSRAQESLQGGWLLVVPQGRAGDVRVTQSVAEVAAKGGEATVL
EVDALHPDRAAYAEALTRWPGVRGVVSFLAWEEQALAEHPVLSAGLAASLALAQALID
VGGSGESAPRLWLVTEAAVVIGAADTGAVIDPVHAQLWGFGRVLALEHPELWGGLIDL
PAVAGEPGSITDHAHADLLATVLATMVQAAARGEDQVAVRTTGTYVPRLVRSGGSAHS
GARRWQPRDTVLVTGGMGPLTAHIVRWLADNGADQVVLLGGQGADGEAEALRAEFDGH
TTKIELADVDTEDSDALRSLLDRTTGEHPLRAVIHAPTVVEFASVAESDLVRFARTIS
SKIAGVEQLDEVLSGIDTAHDVVFFSSVAGVWGSAGQSAYAAGNAFLDAVAQHRRLRG
LPGTSVAWTPWDDDRSLASLGDSYLDRRGLRALSIPGALASLQEVLDQDEVHAVVADV
DWERFYAGFSAVRRTSFFDDVHDAHRPALSTAATNDGQARDEDGGTELVRRLRPLTET
EQQRELVSLVQSEVAAVLGHSSTDAVQPQRAFREIGFDSLTAVQLRNRLTATTGMRLP
TTLVFDYPTTNGLAEYLRSELFGVSGAPADLSVVRNADEEDDPVVIVGMACRFPGGID
TPEAFWKLLEAGGDVISELPANRGWDMERLLNPDPEAKGTSATRYGGFLYDAGEFDAA
FFGISPREALAMDPQQRLLLETVWELIESAGVAPDSLHRSRTGTFIGSNGQFYAPLLW
NSGGDLEGYQGVGNAGSVMSGRVAYSLGLEGPAVTVDTACSSSLVALHLAVQALRRGE
CSLAIAGGVTVMSTPDSFVEFSRQQGLSEDGRCKAFASTADGFGLAEGVSALLVERLS
DARRLGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRAALADAGLAPADVDVVEA
HGTGTRLGDPIEAQALLATYGQGRAGGRPVVLGSVKSNIGHTQAAAGVAGVMKMVLAL
ERGVVPKTLHVDEPSPHVDWSAGEVELAVEAVPWSRGGRVRRAGVSSFGISGTNAHVI
VEEAPAEPEPEPGTRVVAAGDLVVPWVVSGRDAGALREQAARLAAHVSSTGAGVVDVG
WSLVATRSVFEHRAVMVGSELDSMAESLAGFAAGGVVPGVVSGVAPAEGRRVVFVFPG
QGSQWVGMAAGLLDACPVFAEAVAECAAVLDPVTGWSLVEVLRGGGEAVLGRVDVVQP
ALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSRAIAR
IAGGGGMVSVGLSAERVRTMLDTYGGRVSVAAVNGPSSTVVSGDVQALDELLAGCERE
GVRARRVPVDYASHSAQMDQLRDELLEALADITPQHSSVPFFSTVTADWLDTTALDAG
YWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDALDQNAAVLGSLR
RDEGGLDRLLTSLAEAFVQGVPVDWTHAFEGMTPRTVDLPTYPFQRQHYWPKPAPAPG
ANLGDVASVGLTAAGHPLLGAVVEMPDSDGLVLTGQISLRTHPWLADHEVLGSVLLPG
TAFVELAVQAADRAGYDVLDELTLEAPLVLPDRGGIQVRLALGPSEADGRRSLQLHSR
PEEAAGFHRWTRHASGFVVPGGTGAARPTEPAGVWPPAGAEPVALASDRYARLVERGY
TYGPSFQGLHTAWRHGDDVYAEVALPEGTPADGYALHPALLDAAVQAVGLGSFVEDPG
QVYLPFLWSDVTLHATGATSLRVRVSPAGPDTVALALADPAGAPVATVGALRLRTTSA
```

```
  Query Match             44.8%;  Score 110.2;  DB 1;  Length 24568;
  Best Local Similarity   66.8%;  Pred. No. 1e-13;
  Matches  157;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy         12 GACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCATGTGGCGGCAGT 71
              | || |  | |||| |||     ||  ||| ||| ||||| || || ||    ||
Db      24001 GGCGGCACCGGACCCACGTCGTGCCCTGCGCGACCTCGTCCGCGGCCACGTCGCCCTGGT 24060

Qy         72 GCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGTGCACTCGGCTT 131
              |||| | |||   | |||  | |||||||    |||| ||| | ||||| |    |||| ||
Db      24061 GCTCGGACACAGTGGCCCCGAGGCCATCGACGCCGAACAGGCCTTCCGGGACATCGGTTT 24120

Qy        132 CGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACAGGACTCCGCCT 191
              ||| || || || || |||||  || |||||| ||||| |   |  || || |||||| |
Db      24121 CGACTCCCTGACCGCAGTCGAACTCAGAAACCGGCTGAACGCCGAGACCGGCCTCCGCTT 24180

Qy        192 TCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTACACCTG 246
              || | | |||| || |||||| |||| | |||   |   |||||||  ||||||
Db      24181 GCCCGGCACGCTCGTGTTCGACTACCCCAACCCGAGCGCGCTCGCCGATCACCTG 24235


RESULT 14
AR159871
LOCUS       AR159871               50937 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 1 from patent US 6251636.
ACCESSION   AR159871
VERSION     AR159871.1  GI:16222696
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 50937)
  AUTHORS   Betlach,M.C., Shah,S.Krishnakant., McDaniel,R. and Tang,L.
  TITLE     Recombinant oleandolide polyketide synthase
  JOURNAL   Patent: US 6251636-A 1 26-JUN-2001;
FEATURES             Location/Qualifiers
     source          1. .50937
                     /organism="unknown"
BASE COUNT     6672 a  16252 c  19273 g   8740 t
ORIGIN

  Query Match             44.8%;  Score 110.2;  DB 6;  Length 50937;
  Best Local Similarity   66.8%;  Pred. No. 8.8e-14;
  Matches  157;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy         12 GACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCATGTGGCGGCAGT 71
              | || |  | |||| |||     ||  ||| ||| ||||| || | ||| || ||    ||
Db      29150 GGCGGCACCGGACCCACGTCGTGCCCTGCGCGACCTCGTCCGCGGCCACGTCGCCCTGGT 29209

Qy         72 GCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGTGCACTCGGCTT 131
              |||| | |||   | |||  | |||||||    |||| ||| | ||||| |    |||| ||
Db      29210 GCTCGGACACAGTGGCCCCGAGGCCATCGACGCCGAACAGGCCTTCCGGGACATCGGTTT 29269

Qy        132 CGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACAGGACTCCGCCT 191
              ||| || || || || |||||  || |||||| ||||| |   |  || || |||||| |
Db      29270 CGACTCCCTGACCGCAGTCGAACTCAGAAACCGGCTGAACGCCGAGACCGGCCTCCGCTT 29329

Qy        192 TCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTACACCTG 246
              || | | |||| || |||||| |||| | |||   |   |||||||  ||||||
```

```
     misc_feature    5878. .11379
                     /standard_name="module 6 (approximate)"
                     /note="approximate domain spans: ketoacyl synthase, bp
                     5878. .7164; acyltransferase, bp 7477. .8478; ketoreductase,
                     bp 9382. .9930; acyl carrier protein, bp 10213. .10467;
                     thioesterase, bp 10636. .11379; putative"
BASE COUNT     1387 a   3167 c   4610 g   2316 t
ORIGIN

  Query Match           44.8%;  Score 110.2;  DB 1;  Length 11480;
  Best Local Similarity   66.8%;  Pred. No. 1.2e-13;
  Matches  157;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy          12 GACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCATGTGGCGGCAGT 71
               | || |  | |||| |||    ||  ||| ||| ||||| || | ||| || ||    ||
Db         183.GGCGGCACCGGACCCACGTCGTGCCCTGCGCGACCTCGTCCGCGGCCACGTCGCCCTGGT 242

Qy          72 GCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGTGCACTCGGCTT 131
               |||| | |||   | |||  | |||||||   |||| ||| | ||||| |   |||| ||
Db         243 GCTCGGACACAGTGGCCCCGAGGCCATCGACGCCGAACAGGCCTTCCGGGACATCGGTTT 302

Qy         132 CGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACAGGACTCCGCCT 191
               ||| || || || || ||||  || |||||| ||||| |   |  || || |||||| |
Db         303 CGACTCCCTGACCGCAGTCGAACTCAGAAACCGGCTGAACGCCGAGACCGGCCTCCGCTT 362

Qy         192 TCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTACACCTG 246
                 || | | |||| || |||||| |||| | |||   |  |||||||  ||||||
Db         363 GCCCGGCACGCTCGTGTTCGACTACCCCAACCCGAGCGCGCTCGCCGATCACCTG 417


RESULT 13
AF220951
LOCUS       AF220951               24568 bp    DNA     linear   BCT 12-JUL-2000
DEFINITION  Streptomyces antibioticus 8,8a-deoxyoleandolide synthase 1 (oleAI)
            and 8,8a-deoxyoleandolide synthase 2 (oleAII) genes, complete cds.
ACCESSION   AF220951
VERSION     AF220951.1  GI:9049534
KEYWORDS    .
SOURCE      Streptomyces antibioticus.
  ORGANISM  Streptomyces antibioticus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 24568)
  AUTHORS   Shah,S., Xue,Q., Tang,L., Carney,J.R., Betlach,M. and McDaniel,R.
  TITLE     Cloning, characterization and heterologous expression of a
            polyketide synthase and P-450 oxidase involved in the biosynthesis
            of the antibiotic oleandomycin
  JOURNAL   J. Antibiot. 53 (5), 502-508 (2000)
  MEDLINE   20363406
   PUBMED   10908114
REFERENCE   2  (bases 1 to 24568)
  AUTHORS   Shah,S., Betlach,M. and McDaniel,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JAN-2000) Kosan Biosciences, Inc., 3832 Bay Center
            Place, Hayward, CA 94545, USA
FEATURES             Location/Qualifiers
     source          1. .24568
                     /organism="Streptomyces antibioticus"
                     /db_xref="taxon:1890"
     gene            623. .13075
                     /gene="oleAI"
     CDS             623. .13075
                     /gene="oleAI"
                     /note="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="8,8a-deoxyoleandolide synthase 1"
                     /protein_id="AAF82408.1"
                     /db_xref="GI:9049535"
                     /translation="MHVPGEENGHSIAIVGIACRLPGSATPQEFWRLLADSADALDEP
                     PAGRFPTGSLSSPPAPRGGFLDSIDTFDADFFNISPREAGVLDPQQRLALELGWEALE
                     DAGIVPRHLRGTRTSVFMGAMWDDYAHLAHARGEAALTRHSLTGTHRGMIANRLSYAL
                     GLQGPSLTVDTGQSSSLAAVHMACESLARGESDLALVGGVNLVLDPAGTTGVERFGAL
                     SPDGRCYTFDSRANGYARGEGGVVVVLKPTHRALADGDTVYCEILGSALNNDGATEGL
                     TVPSARAQADVLRQAWERARVAPTDVQYVELHGTGTPAGDPVEAEGLGTALGTARPAE
                     APLLVGSVKTNIGHLEGAAGIAGLLKTVLSIKNRHLPASLNFTSPNPRIDLDALRLRV
                     HTAYGPWPSPDRPLVAGVSSFGMGGTNCHVVLSELRNAGGDGAGKGPYTGTEDRLGAT
                     EAEKRPDPATGNGPDPAQDTHRYPPLILSARSDAALRAQAERLRHHLEHSPGQRLRDT
                     AYSLATRRQVFERHAVVTGHDREDLLNGLRDLENGLPAPQVLLGRTPTPEPGGLAFLF
                     SGQGSQQPGMGKRLHQVFPGFRDALDEVCAELDTHLGRLLGPEAGPPLRDVMFAERGT
                     AHSALLSETHYTQAALFALETALFRLLVQWGLKPDHLAGHSVGEIAAAHAAGILDLSD
                     AAELVATRGALMRSLPGGGVMLSVQAPESEVAPLLLGREAHVGLAAVNGPDAVVVSGE
                     RGHVAAIEQILRDRGRKSRYLRVSHAFHSPLMEPVLEEFAEAVAGLTFRAPTTPLVSN
                     LTGAPVDDRTMATPAYWVRHVREAVRFGDGIRALGKLGTGSFLEVGPDGVLTAMARAC
                     VTAAPEPGHRGEQGADADAHTALLLPALRRGRDEARSLTEAVARLHLHGVPMDWTSVL
                     GGDVSRVPLPTYAFQRESHWLPSGEAHPRPADDTESGTGRTEASPPRPHDVLHLVRSH
                     AAAVLGHSRAERIDPDRAFRDLGFDSLTALELRDRLDTALGLRLPSSVLFDHPSPGAL
                     ARFLQGDDTRRPEPGKTNGTRATEPGPDPDDEPIAIVGMACRFPGGVTSPEDLWRLLA
                     AGEDAVSGFPTDRGWNVTDSATRRGGFLYDAGEFDAAFFGISPREALVMDPQQRLLLE
                     TSWEALERAGVSPGSLRGSDTAVYIGATAQDYGPRLHESDDDSGGYVLTGNTASVASG
                     RIAYSLGLEGPAVTVDTACSSSLVALHLAVQALRRGECSLALAGGATVMPSPGMFVEF
                     SRQGGLSEDGRCKAFAATADGTGWAEGVGVLLVERLSDARRLGHRVLAVVRGSAVNQD
                     GASNGLTAPNGPSQQRVIRAALADAGLVPADVDVVEAHGTGTRLGDPIEAQALLATYG
                     QGRAGGRPVVLGSVKSNIGHTQAAAGVAGVMKMVLALGRGVVPKTLHVDEPSAHVDWS
                     AGEVELAVEAVPWSRGGRVRRAGVSSFGISGTNAHVIVEEAPAEPEPEPERGPGSVVG
                     VVPWVVSGRDAGALREQAARLAAHVSGVSAVDVGWSLVATRSVFEHRAVMVGSELDAM
                     AESLAGFAAGGVVPGVVSGVAPAEGRRVVFVFPGQGSQWVGMAAGLLDACPVFAEAVA
                     ECAAVLDPLTGWSLVEVLRGGGEAVLGRVDVVQPALWAVMVSLARTWRYYGVEPAAVV
                     GHSQGEIAAACVAGGLSLADGARVVVLRSRAIARIAGGGGMVSVSLPAGRVRTMLEEF
                     DGRVSVAAVNGPSSTVVSGDVQALDELLAGCEREGVRARRVPVDYASHSAQMDQLRDD
                     LLEALATIVPTSANVPFFSTVTADWLDTTALDAGYWFTNLRETVRFQEAVEGLVAQGM
                     GAFVECSPHPVLVPGITETLDTFDADAVALSSLRRDEGGLDRFLTSLAEAFVQGVPVD
                     WSRAFEGASPRTVDLPTYPFQRQRYWLLDKAAQRERERLEDWRYHVEWRPVTTRPSAR
                     LSGVWAVAIPARLARDSLLVGAIDALERGGARAVPVVVDERDHDRQALVEALRNGLGD
                     DDLAGVLSLLALDEAPHGDHPDVPVGMAASLALVQAMADAAAEVPVWFATRGAVAALP
                     GESPERPRQALLWGLGRVVALEQPQIWGGLVDLPQHLDEDAGRRLVDVVGGLADEDQL
                     AVRASSVLARRLVRTPGHRMSSQAGGREWSPSGTVLVTGGTGALGAHVARWLAGKGAE
                     HLVLISRRGADAAGAAALRDSLTDMGVRVTLAACDAADRHALETLLDSLRTDPAQLTA
                     VIHAAGALDDGMTTVLTPEQMNNALRAKVTATVNLHELTRDLDLSAFVLFSSISATLG
                     IPGQANYAPGNSFLDAFAEWRRAQGLVATSIAWGPWSGGTGMAHEGSVGERLQRHGVL
                     AMEPAAAIAALDHTLASDETAVAVADIDWSRFFLAYTALRARPLIGEIPEARRMLESG
                     SGPGDLEPDRAEPELAVRLAGLTAVEQERLLVQLVREQAAVVLGHSGAEAVAPDRAFK
                     DLGFDSLTSVELRNRLNTATGLRLPVTAVFDYARPAALAGHLRSRLIDDDGDHGALPG
                     VEKHAIDEPIAIVGMACRFPGGIASPEDLWDVLTAGEDVVSGLPQNRGWDLGRLYDPD
                     PDRAGTSYMREGAFLHEAGEFDAAFFGISPREALAMDPQQRLLLETSWEALERAGITP
                     SKLAGSPTGVFFGMSNQDYAAQAGDVPSELEGYLLTGSISSVASGRVAYTFGLEGPAV
                     TVDTACSSSLVALHLAVQGLRRGECSLALVGGVTVMSSPVTLTTFSRQRGLSVDGRCK
                     AFAASADGFGAAEGVGVLLVERLSDARRLGHRVLAVVRGSAVNQDGASNGLAAPNGPS
                     QQRVIRAALADAGLAPADVDVVEAHGTGTRLGDPIEAQALLATYGQQRTSGRPVWLGS
                     VKSNIGHTQAAAGVAGVMKMVLALGRGVVPKTLHVDEPSPHVDWSAGEVELAVEAVPW
                     SRGGRVRRAGVSSFGISGTNAHVIVEEAPAEPSVEEGPGSVVGVVPWVVSGRDAGALR
                     AQAARLAAHVSSTGAGVVDVGWSLVATRSVFEHRAVMVGTDLDSMAGSLAGFAAGGVV
                     PGVVSGVAPAEGRRVVFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDRLTGWS
                     LVEVLRGGEAVLGRVDVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAG
                     GLSLADGARVVVLRSRAIARIAGGGGMVSVGLSAERVRTMLDTYGGRVSVAAVNGPSS
                     TVVSGDAQALDELLAGCEREGVRARRVPVDYASHSAQMDQLRDELLEALADVTPQDSS
                     VPFFSTVTADWLDTTALDAGYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVP
                     GITETLDTFDADAVALSSLRRDEGGLDRFLTSLAEAFVQGVPVDWTHAFEGGRPRFVD
                     LPTYAFQRQRYWLHEEPLQEPVDEAWDAEFWSVVERGDATAVSDLLSTDAEALHTVLP
                     ALSSWRRRRVEHRRLQDWRYRVEWKPFPAALDEVLGGGWLFVVPRGLADDGVVARVVA
                     AVTARGGEVSVVELDPTRPDRRAYAEAVAGRGVSGVVSFLSWDDRRHSEHSVVPAGLA
                     ASLVLAQALVDLGRVGEGPRLWLVTRGAVVAGPSDAGVVIDPVQAQVWGFGRVLGLEH
                     PELWGGLVDLPVGVDEEVCRRFVGVVASAGFEDQVAVRGSGVWVRRLVRAVVDGGGGG
                     WRPRGTVLVTGGLGGLGAHTARWLVGGGADHVVLVSRRGGSAPGAGDLVRELEGLGGA
                     RVSVRACDVADRVALRALLSDLGEPVTAVFHAAGVPQSTPLAEISVQEAADVMAAKVA
                     GAVNLGELVDPCGLEAFVLFSSNAGVWGSGGQAVYAAANAFLDALAVRRRGVGLPATS
                     VAWGMWAGEGMASVGGAARELSRRGVRAMDPERAVAVMADAVGRGEAFVAVADVDWER
                     FVTGFASARPRPLISDLPEVRAVVEGQVQGRGQGLGLVGEEESSGWLKRLSGLSRVRQ
                     EEELVELVRAQAAVVLGHGSAQDVPAERAFKELGFDSLTAVELRNGLAAATGIRLPAT
                     MAFDHPTATAIARFLQSELVGSDDPLTLMRSAIDQLETGLALLESDEEARSEITKRLN
                     ILLPRFGSGGSSRGREAGQDAGEHQDVEDATIDELFEVLDNELGNS"
     misc_feature    650. .1906
                     /gene="oleAI"
                     /note="Region: ketoacyl synthase domain KS-Q"
     misc_feature    2309. .3415
                     /gene="oleAI"
```

```
                     /protein_id="BAB69304.1"
                     /db_xref="GI:15824140"
                     /translation="MSCRYPGGVDSPEDLWELVAGGRDAVSEFPADRGWDLDALYDPD
                     PESRGTSYAREGAFLRDLARFDAGLFGISPREALAMDPQQRLLLETSWELFERAGIDA
                     ASLAGSRTGVFAGVMNHEFLAALQNSPQDLEGYLGTGTSGSVASGRVAYTFGLEGPAV
                     TVDTACSSSLVALHLAVQSLRNGECSLAVAGGVTAMVGPATFVGFSRQRGLAPDGRCK
                     PFAAGADGTGWGEGAGLLLVERLSDARRNGHEVLAVVRGSAVNQDGASNGLTAPNGPS
                     QQRVIRQALAAAQLTTAQVDAVEAHGTGTRLGDPIEAQALLATYGQGRPDDKPLWLGS
                     IKSNIGHTQAAAGVAGIIKMVMAMRHGELPRTLHVDAPSVHVDWTAGAVRLLTERTPW
                     SASDQPRRAGISSFGISGTNAHTIIEEAPSAEAFAQDTGAPEGPEEGAESAEPVASEA
                     ARPAAGGPVPWVISGRTEEALRAQAERLLSRTDAEVPPLDVAFSLATTRTALEHRAVL
                     QGETPAELAEGLRAVASGSPARGVARGLSGTGGRVGFLFSGQGSQRLGMGRELYASYP
                     AFAAAYDEVCAQLDAVVDVDSEELHRTGAAQPALFAVEVALFRLLESWGIRADYVAGH
                     SVGELAAAHVAGVLSLDDAAKLVSARARLMQALPAGGAMVAVQATEDEVLPHLTDAVG
                     IAAINGPRSVVVSGAEDAVTAIAEAFRQQGRKSSRLKVSHAFHSPLMDPILEEFAEVV
                     RSLTFHRPQIPVVSNLTGRLAEPYTPQYWVRHVREAVRFADGVRTLNDLGVTTFVEIG
                     PGGVLSALAEGCLDDDVVTVPALRADRPERVALVAAVSELHAHGVSPDWHAFFPGARR
                     VDLPTYAFQYERYWLDASPAAPGGVRAAGLGSADHPLLGAAVSLAGGDERLLTGRLSL
                     RTHPWLSDHAVLGTALLPGTAFVELAVRAADEAGCALLEDLTLEAPLVVPEQGGVAVQ
                     VWVGAVDGTGRRPLTVHSRPEDDSDLPWVRHATGFVTEGSEEAGAAAADRTLTAWPPT

  Query Match          45.0%;  Score 110.8;  DB 1;  Length 78210;
  Best Local Similarity  67.1%;  Pred. No. 6e-14;
  Matches  157;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy         13 ACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCATGTGGCGGCAGTG 72
              ||| |||  || ||  |   |.   |||| || || || || || || |  ||  | |||
Db      10282 ACGCCGGGGGAGCAGGGCCGCATCCTGCTGGACCTGGTCCGCGAGCAGGCCGCCACGGTG 10341

Qy         73 CTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGTGCACTCGGCTTC 132
              ||| | ||| ||| |  |   ||| |||   |||||| | ||||||| |    |||||||
Db      10342 CTCGGCCACACGGGCAAGGGCGCCTTCGAAGCCGACCGGACGTTCCGCGAGACCGGCTTC 10401

Qy        133 GATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACAGGACTCCGCCTT 192
              || || ||||| ||||||||  | ||  ||  ||| | || ||| || || ||| | ||
Db      10402 GACTCGCTCACCGCCGTCGAACTGCGTCACAGGCTCAACACGGCCACTGGCCTCAGGCTG 10461

Qy        193 CCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTACACCTG 246
              ||   | | ||||||||||||||||||||||  ||   ||||||   ||||||
Db      10462 CCCACCACCCTGGTCTTCGACCACCCGACCCCAACCGCCCTCGCCCGTCACCTG 10515


RESULT 12
STMPKS3ORF
LOCUS       STMPKS3ORF             11480 bp    DNA     linear   BCT 14-JUL-1994
DEFINITION  Streptomyces antibioticus polyketide synthase gene, complete cds of
            ORF3 subunit including module 5 and module 6.
ACCESSION   L09654
VERSION     L09654.1  GI:153407
KEYWORDS    acyl carrier protein; acyltransferase; deoxyoleandolide; ketoacyl
            synthase; ketoreductase; lactone ring; module; oleandomycin;
            polyketide synthase; thioesterase.
SOURCE      Streptomyces antibioticus (library: ATCC 11891) DNA.
  ORGANISM  Streptomyces antibioticus
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 11480)
  AUTHORS   Swan,D.G., Rodriguez,A.M., Vilches,C., Mendez,C. and Salas,J.A.
  TITLE     Characterisation of a Streptomyces antibioticus gene encoding a
            type I polyketide synthase which has an unusual coding sequence
  JOURNAL   Mol. Gen. Genet. 242 (3), 358-362 (1994)
  MEDLINE   94150470
   PUBMED   8107683
FEATURES             Location/Qualifiers
     source          1. .11480
                     /organism="Streptomyces antibioticus"
                     /db_xref="taxon:1890"
     CDS             820. .11379
                     /function="formation of 8,8-deoxyoleandolide; biosynthesis
                     of oleandomycin macrolactone ring"
                     /standard_name="PKS; OLE-ORF3"
                     /note="subunit orf3, including module 5 and module 6;
                     putative"
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAA19695.1"
                     /db_xref="GI:153408"
                     /translation="MAEAEKLREYLWRATTELKEVSDRLRETEERAREPIAIVGMSCR
                     FPGGGDATVNTPEQFWDLLNSGGDGIAGLPEDRGWDLGRLYDPDPDRAGTSYVREGGF
                     LYDSGEFDAAFFGISPREALAMDPQQRLLLETSWEAFESAGIKRAALRGSDTGVYIGA
                     WSTGYAGSPYRLVEGLEGQLAIGTTLGAASGRVAYTFGLEGPAVTVDTACSSSLVALH
                     LAVQGLRRGECSLALVGGVTVMSSPVTLTTFSRQRGLSVDGRCKAFPASADGFGAAEG
                     VGVLLVERLSDARRLGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRAALADAGL
                     APADVDVVEAHGTGTRLGDPIEAQALLATYGQGRAGGRPVWLGSVKSNIGHTQAAAGV
                     AGVMKMVLALGRGVVPKTLHVDEPSPHVDWSAGAVELLTEERPWEPEAERLRRAGISA
                     FGVSGTNAHVIVEEAPAEPEPEPGTRVVAAGDLVVPWVVSGRDARALRAQAARLAAHV
                     SGVSAVDVGWSLVATRSVFEHRAVAIGSELDSMAGSLAGFAAGGVVPGVVSGVAPAEG
                     RRVVFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPVTGWSLVEVLQGRDATV
                     LGRVDVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVV
                     VLRSRAIARIAGGGGMVSVSLPAGRVRTMLEEFDGRLSVAAVNGPSSTVVSGDVQALD
                     ELLAGCEREGVRARRVPVDYASHSAQMDQLRDELLEALADITPQDSSVPFFSTVTADW
                     LGTTALGAGYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDALDQ
                     NAAVFGSLRRDEGGLDRFLTSLAEAFVQGVPVDWSRAFEGVTPRTVDLPTYPFQRQHY
                     WLMAEEAPVSQPPHSENSFWSVVADADAEAAAELLGVDVEAVEAVMPALSSWHRQSQL
                     RAEVNQWRYDVAWKRLTTGALPEKPGNWLVVTPAGTDTTFAESLARTAAAELGVSVSF
                     AQVDTAHPDRSQYAHALRQALTGPENVDHLVSLLALDQATDDLAAAPSCLAASLVLAQ
                     ALVDLGRVGEGPRLWLVTRGAVVAGPSDAGAVIDPVQAQVWGFGRVLGLEHPELWGGL
                     IDLPVGVDEEVCRRFVGVVASAGFEDQVAVRGSGVWVRRLVRAVVDGGGGGWRPRGTV
                     LVTGGLGGLGAHTARWLVGGGADHVVLVSRRGGSAPGAGDLVRELEGLGGARVSVRAC
                     DVADRVALRALLSDLGEPVTAVFHAAGVPQSTPLAEISVQEAADVMAAKVAGAVNLGE
                     LVDPCGLEAFVLFSSNAGVWGSGGQAVYAAANAFLDALAVRRRGVGLPATSVAWGMWA
                     GEGMASVGGAARELSRRGVRAMDPERAVAVMADAVGRGEAFVAVADVDWERFVTGFAS
                     ARPRPLISDLPEVRAVVEGQVQGRGQGLGLVGEEESSGWLKRLSGLSRVRQEEELVEL
                     VRAQAAVVLGHGSAQDVPAERAFKELGFDSLTAVELRNGLAAATGIRLPATMAFDHPN
                     ATAIARFLQSQLLPDAESESAVPSSPEDEVRQALASLSLDQLKGAGLLDPLLALTRLR
                     EINSTVQNPEPTTESIDEMDGETCCAWRSAKSTAEPLTTGADMPDPTAKYVEALRASL
                     KENERLRQQNHSLLAASREAIAITAMSCRFGGGIDSPEDLWRFLAEGRDAVAGLPEDR
                     GWDLDALYHPDPENPGTTYVREGAFRYDAAQFDAGFFGISPREALAMDPQQRLLLETS
                     WELFERADIDPYTVRGTATGIFIGAGHQGYGPDPKRAPESVAGYLLTGTASAVLSGRI
                     SYTFGLEGPAVTVDTACSSSLVALHLAVQALRRGECSLAIAGGVAVMSTPDAFVEFSR
                     QQGMARDGRCKAFAAAADGMGWGEGVSLLLLERLSDARRLGHRVLAVVRGSAVNQDGA
                     SNGLAAPNGPSQQRVIRAALADAGLAPADVDVVEAHGTGTRLGDPIEAQALLATYGQG
                     RAGGRPVWLGSVKSNIGHTQAAAGVAGVMKMVLALGRGVVPKTLHVDEPSPHVDWSAG
                     AVELLTEERPWEPEAERLRRAGISAFGVSGTNAHVIVEEAPAEPEPEPGTRVVAAGDL
                     VVPWVVSGRDVGALREQAARLAAHVSSTGAGVVDVGWSLVATRSVFEHRAVMVGTDLD
                     SMAGSLAGFAAGGVVPGVVSGVAPAEGRRVVFVFPGQGSQWVGMAAGLLDACPVFAEA
                     VAECAAVLDPVTGWSLVEVLQGRDATVLGRVDVVQPALWAVMVSLARTWRYYGVEPAA
                     VVGHSQGEIAAACVAGGLSLADGARVVVLRSRAIARIAGGGGMVSVSLPAGRVRTMLD
                     TYGGRVSVAAVNGPSSTVVSGDVQALDELLAGCEREGVRARRVPVDYASHSAQMDQLR
                     DELLEALADITPQDSSVPFFSTVTADWLDTTALDAGYWFTNLRETVRFQEAVEGLVAQ
                     GMGAFVECSPHPVLVPGIEQTLDALDQNAAVLGSLRRDEGGLDRLLTSLAEAFVQGVP
                     VDWTHAFEGVTPRTVDLPTYPFQRQRFWLDGSPASSANGVDGEADAMIWDAVEREDSV
                     AVAEELGIDAEALHTVLPALSSWRRRRVEHRRLQDWRYRVEWKPFPAALDEVLGGGWL
                     FVVPRGLADDGVVARVVAAVTARGGEVSVVELDPTRPDRRAYAEAVAGRGVSGVVSFL
                     SWDDRRHSEHPVVPAGLAASLVLAQALVDLGRVGEGPRLWLVTRDAVVAGPSDAGAVI
                     DPVQAQVWGFGRVLGLEHPELWGGLIDLPVEAPEPGSTCDHTYADLLATVVASAGFED
                     QVAVRGSGVWVRRLRAVVDGGGGGWRPRGTVLVTGGLGGLGAHTARWLVGGGADHVV
                     LVSRRGGSAPGAGDLVRELEGLGGARVSVRACDVADRVALRALLSDLGEPVTAVFHAA
                     GVPQSTPLAEISVQEAADVMAAKVAGAVNLGELVDPCGLEAFVLFSSNAGVWGSGGQA
                     VYAAANAFLDALAVRRRGVGLPATSVAWGMWAGEGMASVGGAARELSRRGVRAMDPER
                     AVAVMADAVGRGEAFVAVADVDWERFVTGFASARPRPLISDLPEVRTALRNQEQEQLH
                     APVPEDRSAQLLRRLSMLSPAGREAELVKLVRTEAAAVLGHGSAQDVPAERAFKELGF
                     DSLTAVQLRNRLAAATGTRLPASAVFDHPHAAALARWLLAGMRHADGGHGGGHAGGPG
                     PDADEGRSAGAGHSGMLADLYRRSAELGRSREFIGLLADTAAFRPVFHGPADLDAPLE
                     AVPLADGVRKPQLICCSGTAPVGGPHEFARLASFFRGTRAVSALPLPGYLPGEQLPAD
                     LDAVLAAQAEAVEKQTGGAPFVLVGYSAGGLMAHALACHLAGRGTPPSGEVLVDVYPP
                     GRQEPVFGWQKELTEGMFAQDFVPMDDTRLTALGTYDRLMGEWRPAPSGLPTLLIRAT
                     EPMAEWTGAIDWRASWEYDHTAVDMPGNHFTIMREHAEDAARHIDVWLKGLTP"
     misc_feature    820. .5505
                     /standard_name="module 5 (approximate)"
                     /note="approximate domain spans: ketoacyl synthase, bp
                     919. .2217; acyltransferase, bp 2524. .3525; ketoreductase,
                     bp 4414. .4962; acyl carrier protein, 5251. .5505; putative"
                     /citation=[1]
```

```
                     /transl_table=11
                     /product="putative lipoprotein"
                     /protein_id="BAB69301.1"
                     /db_xref="GI:15824137"
                     /translation="MGEALITLCTRSTRLAATAGVLGGVLAVTALGGTSNAATSGNGA
                     GNHPTSQARINQAGAQEASDARIEITPGQGTYSVGTNHPVSVTVTSGKLTKVTMTAVA
                     TGAEIPGTLSTDGTSWKPNGRLERATKYQIAAEAGDVKGRSTTGSATITTVSPANDFI
                     GHLSPEDGSTVGVAMPVTVNFDKAISDKAAVQSEIRVSSSSGQRVVGHWFNDHRVDFR
                     PENYWRPGSTVTVTLNRHGIQKTVTFKIGRSQISTVDAKTKQMTVVRDGKTIKTIPIS
                     SGSPEHPTYNGQMVISEKFRQIHMNGASVGLTEKNGKPSYDIKAVPHAMRLTDSGTFI
                     HGNYWGTDSVFGEANTSHGCVGLKDVRGDSHGRQPAAWFFDHSIVGDVVIVKNSEEKT
                     MLAPDNGLSDWNMSWSEWVAGSPAT"
CDS                  complement(1256. .2290)
                     /codon_start=1
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="BAB69302.1"
                     /db_xref="GI:15824138"
                     /translation="MPVEFLSHDQVSRYGRFASAPTPGELEQFFRLDAKALESARSKR
                     APANRLGWAVQWGCVRMLGVFPTEDLSVVPEVVVRFAAEQLGVDPGEFAAYGARRQNR
                     YEHAWEIRDAYAYRESPAAEAQVREFLATRVWASWEGPRALFDRAVGWLVDNRVLLPG
                     ITTLTRLVAEVRAQENAALYRTLDEAVPEKLRQSMRDLLKVPEGRRVSELERLRRSPM
                     GVSGSADRLAFSERIRMEAVAMFTAGPGSTDIAKELRVSVQSVQRWRSGLAGDRPGRG
                     AVPQPVSVELSETLFAALDEERAKGFAPSWTIRYLHHDFPIFGQRYAYKWSQGCLHSF
                     GRSAQGYRDQ"
gene                 2935. .21081
                     /gene="pteA1"
CDS                  2935. .21081
                     /gene="pteA1"
                     /codon_start=1
                     /transl_table=11
                     /product="modular polyketide synthase"
                     /protein_id="BAB69303.1"
                     /db_xref="GI:15824139"
                     /translation="MVESELNATGNRDCSASADSADGIAVIGLACRLPGAADPAAFWR
                     LLSEGADAITDVPPDRWDGAVVADASAPDRTGIRRGGFLDSIGDFDAGFFGLSPKEAA
                     AMDPQQRLVLELAWEALEDAHVRAATLRSTRTGVFVGAMWDDYATLHHRSGLTAISPH
                     TVTGLHRSIIANRVSYFLGLNGPSLTVDSGQSSSLVSVHLACESLRKGESTIALAGGV
                     NLNIVPESALGAAKFGGLSPDGRCFTFDARANGYVRGEGGGIVVLKPLARAVADGDPV
                     YCVIRGSAVNNDGGGDGLTVPLQSGQEQVLRLAYERAGVNPADVGYVELHGTGTKVGD
                     PIEAAALGAVLGTGREPGRPLRVGSAKTNVGHLEGAAGITGLLKAVLSLNHRELPASL
                     NFGTPNPAIPLDRLNLRVQTEHSEWAAEDGRPLLAGVSSFGMGGTNCHIVLASGAAVE
                     AGEGATAGGGAAPTFWPLSAKTASGLRAQAAALLAHTEAPLGLALPDVGWSLATGRTA
                     FEHRAVVVGEVRDDFLRALRELSTGGIDAALTTGRAGPRGELAFLFSGQGSQRAGMGR
                     ELAESVPAFASALDEVCAHLDPLLPRPLREVMFAPVGTGEAAELDQTLYTQTSLFAVE
                     VALFRLLESWDITPDVLMGHSIGELAAAHVAGVLSLADACALVAARGRLMQALPAGGA
                     MIAVQAAEDEVRARIGERTDRVSIAALNGPDSVVVSGDEDLATEIADAFAALGRKTSR
                     LRVSHAFHSPHMEPMLAEFRNVAEGLTFHAPRIPIVSNVTGRLSEEPSGYEGCTAAYW
                     VRHVREAVRFADGVARLDEQGVRTYLELGPGGVLASMARAGADSESLFVPALRTRRPE
                     AQALLAAVASLHVHGVEPDWDALFGGGRGARRKVALPTYAFERQRYWLDGDTTSPVAT
                     TGWAPVAPQPDAIAETDSAPVGALGRRLAGLAESERDEAVLDMVRAHIAAVLGHAEAR
                     QVETEWTFKDLGFDSLSSVELRNQLSEATGLRLPSGLLFDHPTPAALARHLAAQSLGG
                     GDAAAAAPATVADPHEPIAIVAMSCRLPGGVSSPEDLWRLLASGGDGISGFPTDRGWD
                     VETLYDPEPGTPGKSSTRHGGFLHEAAEFDAGFFGISPREAAAIDPQQRLVLETAWEA
                     FERAGIDPATLRGSRAGVFIGATAQDYGPRLHEPADGVEGYLLTGTTTSVASGRVAYS
                     FGLEGPAVTVDTACSSSLVALHLAVQSLRSGECTMALAGGVTVMSTPGMFVEFSRQRG
                     LSADGRCKAFAAGADGTGWAEGVGMLLVERLSDARRNGHQVLAVVRGSAVNQDGASNG
                     LTAPNGPSQQRVVRQALAAAGLAPADVDAVEAHGTGTLLGDPIEAEALIAAYGQGRDG
                     DRPLWLGSLKSNIGHTQAAAGVAGVIKMVMAMRHGVLPRTLHVDEPTPHVDWSASGVR
                     LLTEAVEWPARDLPRRAAVSSFGISGTNAHVIVEQAATSEEPARGTAVAGPVPWVVSA
                     KSDVALREQAGRLLSFLDAEGVGGAGVRPEDVALSLATSRAVLERRAAVVGGDLAELR
                     QGLEALASGASGAAPVGGAVGGRVGFLFSGQGSQRIGMGRELYAAYPVFAAAYDEVCA
                     FLDAPVDVDSEELNQTGTTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVA
                     GVLSLEDAAKLVSARAALMQALPDGGAMVAVQATEDEVLRHLTDGVGIAAVNGPRSVV
                     VSGAEDAVTAVGEIFAGRGRKTSRLKVSHAFHSPLMDPMLEEFAAVVRGLTFSESRIP
                     VVSNLTGRLAEVYTPEYWVRHVREAVRFADGIETLGELGVTTFVEVGPGGVLSALAQG
                     CFDDGTDIVTVPVLRADRPEPYALVTAIGRLHTHGISPDWQVLFPGARRVDLPTYAFQ
                     RERYWLDAPRAATAGAPDAMDAEFWDSVESEDRASLGALLGLEPTELDVVASKLSAWR
                     RKRREQSTTDGWRYRITWQPLDDMVRAAPSGTWLYVVSEETAWTEAIRAGLTELGVAL
                     VPFAVGEDTDRDALARAFADAGQEQADGILFAAAQGEAGANALQRLVLLVQALGDAGI
                     EAPLWCLTSGAVSTGTADLLTDPVSAQLWGLGRIVALEQPQRWGGLVDLPAELDSRAL
                     ERLAGLLAQADEDQLAVRASGVYGRRLVRAPLSAAPADAPWSPRGTVLVTGGTGALGG
                     QVARWLAGAGAEHLVLTSRRGPDAPGAAELKAELEELGAQVTVVACDAADRDALAELF
                     GRYIVNAVVHTAGVLDDGLVESLTPERLNHVLRPKVDAALHLHELTRDRQDLDAFVLF
                     SSMTGVWGNGGQGAYGAANAFLDALAEHRRAQGVPALAVAWGSWADGGMADGAAGDHL
                     RRRGVRAIAALPAISVLHGALTHGETSMTVADVDWERFVPAFVGTRPCPLLLGVPEAR
                     EALEARARAAQSPEVPVSALVQRLLGATPGEQGRILLDLVREQAATVLGHTGKGAFEA
                     DRTFRETGFDSLTAVELRHRLNTATGLRLPTTLVFDHPTPTALARHLGEELLGRHGLE
                     TAAAQATAAAAADEPLAIVGMSCRLPGGVQSPEDLWELLSSGRDAMSGFPTDRGWDID
                     ALYDPDPDRPGKTYSREGGFIQGADRFDAGLFGISPREALAMDPQQRLLLETAWEAFE
                     RAGIALASVRASRTGVFIGTNGQDYANGLRNAPEEIEGYALTGKAASVVSGRISYTFG
                     LEGPAVTVDTACSSSLVALHLAAQALRSGECSMALVGGVTVMTTPDLFVEFSRQRGLS
                     PDGRCKAFAAGADGTGWGEGVGLLLVERLSDARRNGHQVLALVRGSAVNQDGASNGLT
                     APNGPSQQRVIRQALASAGLTPADVDAVEAHGTGTRLGDPIEAQALLATYGQEHSDDR
                     PVWLGSVKSNIGHTQAAAGVAGVIKMVLAMRHGVLPQTLHVDEPTAHVDWSAGAVRLL
                     TEPVQWPGTTGPRRAAVSSFGIGGTNAHTIIEEAPAAAAAAEPAREHRPVPVPWVLSAK
                     SEVALRAQAERLLAFATDDVSPVDAGFSSATTRSALEHRAAVIGSDPAELRAGLEALA
                     AGEPAANVVAGRAHSADKVGFLFSGQGSQRLGMGRELYAAYPVFAAAYDEVCALLDAT
                     VDVDAETLHQTGSTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVAGVLSL
                     EDAAKLVSARAALMKALPAGGVMAVQAAEDEMLPHLIDRAGIAAINGPRSVVISGAE
                     DAVTAIAEVFAAQGRKTSRLKVSHAPHSPLMDPMLEEFAEVVRGLTFSEPWIPVVSNL
                     TGRLTESYTPEYWVRHVREAVRFADGVLTLHELGVTTFVEIGPGGVLSALVQGCLDDG
                     TDIVTIPALRADRPEPQALTTAYAQLHVSGAEVDWHAFFPGARRVELPTYAFQRERYW
                     LDAPASAGDMRAVGQGEAGHPLLGAAVPLADSVGHLLTGRLSKHTHPWLVDHAVSGVA
                     LLPGTAFVELAITAADAVGCDLLEELTLETPLLVPERGGVALQVHVGADDGSGRRPLT
                     VHSQIDDGDAYAHEDDSERAWVRHASGVLTVATGGPVDGSLTAWPPVGAEPVDVDGFY
                     DRLAGMALDYGPVFQGLRAAWRSGDDFFAEVELPGAERTDAGAYGLHPALFDAALHTV
                     WLGAVEPEAGTGNGLLPFAWSGVRLAAAGASDLRVKVSRAGTGTSTVSLVLADGTGGP
                     VAEVDSLTLRPVPADQLRAGSGSDDAVFGLKWTPVSLPSAPDGMRIETYEDLAALRAT
                     DPPADTEADAAAAADAVVVPCPTGTGADLATRVREVTHAVLELVQWWLAEERPGRLVL
                     VTRGGDLAQSAVCGLVRSAQTENPDRIVLVETSSDAHAGAGSGVDDADLARVLPGLIA
                     SGEPQAAVHADDEVRVPRLVRVAVPDAEPELPGLGSGTVLLTGASGGLGGLFARHLVA
                     EHGVRSLLLVSRRGDAPGAAELTADLTAQGADVTWAACDVADRAAVRALLAGPGQSL
                     SAIIHTAGVLDDGIIGSLTPERLDAVFRPKVDAALNLHDLAGELSGDLSAFVLFSSVA
                     GTLGTPGQANYAAANTFLDALAEHRRAAGQRATSLAWGLWAQDSESAMTGSLDHTDLT
                     RIQRMGLAAIPSSDGLRLFDAALALDRAVVAPLRLDLGGLQGTVPPVLRGLRGTVRGT
                     ARRTAQTAPKGSLGQRLAGLAEAEREQVVLDLVRTEVAAVLGHAGAQSVGADDSFKDI
                     GFDSLTAVELRNRLNSAVGMRLPATLIFDYPNPLTLAHFLTGEAAGGEAATAVPTTLS
                     TATGTADEPIAIVGMACRYPGGVQSPEDLWQLVFSGRDAVSGFPEGRGWDVEQLYDPN
                     PDQWGTSYTREGGFLHDAADFDAEFFGISPREALAMDPQQRLLLETSWEAFERAGIDP
                     ATVRGSRTGVFAGVMYHDYGGRVHTSPPGLEGYLVNGSAGSIASGRVSYTFGLEGPAV
                     TVDTACSSSLVALHLAAQALRSGECSMALVGGVTVMAGPSVFVEFSRQRGLSADGRCK
                     AFGAGADGTGWAEGVGMLLVERLSDARRNGHKVLAVVRGTAVNQDGASNGLTAPNGPS
                     QQRVIRQALANAGLSSDEVDAVEAHGTGTRLGDPIEAQALIATYGQERSEDRPLWLGS
                     LKSNIGHSQAAAGVGGVIKMVMAMQHGVLPQTLHVDEPTPHVDWSAGAVQLLTEAVEW
                     PASDGPRRAAVSSFGVSGTNAHVIVEQAATSEVTARGTAVAGPVPWVLSAKSDAALRE
                     QAGRLLSFLDAEGVGGAGVRPEDVAYSLATSRAVLECRAAVVGEELGEFRQGLEGIVS
                     GAAPVGGAVGGRVGFLFSGQGSQRIGMGRELYAAYPVFAAAYDEVCALLDATVDVDAE
                     TLHQTGSTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVAGVLSLEDAAKL
                     VSARAALMQALPAGGAMVAIQATEAEVLPHLTHGVGLAAINGPRSVVLSGSEDAVLDI
                     AEVFAQQGRKTSRLKVSHAFHSPLMDPMLEEFAEVVRGLTFSEPRIPVVSNLTGRLAE
                     VYIPEYWVRHVREAVRFADGLHTLHELGVTMFVEIGPGGVLSALVQGCLDDGTHIVTV
                     PALRADRPEPYALVTAVGQLHTHGISPEWQALFPGARRVDLPTYAFQRERYWLDAPQA
                     SGDVRAAGLGAADHPLLGAAIATADSDGAVLTGLLSLGTHPWLADHAVNGIVLLPGTA
                     FVELAIRAGDETGCGWIEDLTLELPLVVPERGGVTLQVAVGPEDESGRRQLSVHSRVD
                     DRSWLRHATGVLSSAEAPSAAKLGAWPPAGAEAVDLTAFYEGMAAAGLDYGPVFQGLR
                     SVWRDGDDVFAEVELPDESAAEAAAFALHPALLDAALHALAAGGLVSLDDSPALPFAW
                     SGVVVQAAGAAMVRVKLSRTATGSVTLAVADGAGDPVASVESLSLRAASAEQLPGAVG
                     GDQLFGLEWTRFTLPSATDALMIETVADFDALRESEGTPDVVVVPVVPCPAGSGRETA
                     DRVHSTTSEVLALVQWWLAEERAGRLALVTRPGDLAHAAVWGLVRSAQSENPDRIVLV
                     EAEETDEAVRVLPAALASGEPQFAVRDGEVFVPRLVKATRSATGTATGTPDFGVGPVL
                     LTGASGALGGLVARHLVAEHGVRSLLLLSRRGAEAPGAVELEAELAAWGAEVRWAACD
                     VADREAVSGMLNGLGEQSLSAVVHTAGVLGDGIVASLTPERMREVFRPKVDAVLNLHE
                     CTRDMGLAAFVVFSSVAGMVGSAGQASYAAANSFLDAFSAHRRREGLPAISLAWGVWE
                     QSGAMTDGLVEADRARMARSGVLPLPPEEGLRLFDAALASDEAVLAPVRIDTGALRAG
                     EAPPVLRALVPAAAARRTARAAAAPSASSLAERLEEMPEAERERKTVLDLVRAEVAAGL
                     GHASDRTVRPEHAFQDLGFDSLTAVELRNRLNKATGLRLPATLVFDHPTPALLARHVF
                     MAAAGAAEPALVDSLLADLDRVEQELVTTLAASEARDRILSKLQGLLAIAGEFGKAPA
                     QEGPGTGGSDLESATDDEVFDLLGKEFGIS"
gene                 21273. .31847
                     /gene="pteA2"
CDS                  21273. .31847
                     /gene="pteA2"
                     /codon_start=1
                     /transl_table=11
                     /product="modular polyketide synthase"
```

```
                /gene="olmRII"
CDS             7658. .10486
                /gene="olmRII"
                /codon_start=1
                /transl_table=11
                /product="luxR-family transcriptional regulatory protein"
                /protein_id="BAB69190.1"
                /db_xref="GI:15823973"
                /translation="MLASAGQGEGGVVLVNAGPGVGRTALLDAFLRKATESGARVCAA
                AGSSAETEAEFGVVGQLFPPTGPLGAGFRLARAIGSRDAGPGLDADRLFDALCQELRR
                RPTVIAVDDAQLADAASLRFLRYLVRRLRTAPVMMVLTEPTGPDALPPSFQAELLRHP
                QCRRLGLLPLSRDGVTQVVSAQVGSDRVPRLAADFHAISGGNPLLVHALLADHRAAHR
                LSPVPSVPTVPSVSSASSAASASSAASAASAVSAAPAASAAFAKAALAWAYRDEPVLF
                DVACGFAVLGESATPARVACLVDRGADAVDQVMTALDTAGLLAGGTFRSALVGKALLE
                HVDVETRAELRRRAARLLHGDGAPATVVARHLLAAPVAEPWAGRTLLAAAEKAMQEGR
                EELSLDCLRLAGQSATTEQERAAVVMARVKIGWEIDPRLVSPWLDELVAALRRGDLRS
                GDAAWIVKHLAWQDRFPEAVESLSTLAAAGDAGEPGVEAELHFVRHWLRHSCPPLEAS
                VPGPSNPTRPRTPRLRRVSPASYAGEVLGKVLTEGPDDHTVAMAEEILRGCRYSETVV
                EALEAALLALVYAGRPDRAAFWCDGLLQQAGTRSTGTSVAILSGIRAEIALRQGALPA
                VEQYARSALSAISLPGWGVAVGSPLAALVHSTTASGMSGVAGAWLNQAVPDGMFRTRH
                GLLYTHARGHYHLAVDRPAAALDDFLTCGGLAKEWGMDVPTFLPWRTSAAQAQLALGN
                PGQARALVREQLGRPGGSTARARGVALRLLAATGEPGRRPALLRESVNLLESCGDQVE
                LLRALADRCQALDHVGAPAKARMTARHARNVADACGGDALFRKLFRDEVPDREGDPDG
                LTGEEQGSASLTAAERRVTALAALGHSNREIGRKLFITKSTVEQHLTRVYRKLGVRNR
                ADLGDLLAGTNLGAGTPAMAMASSAVAG"
CDS             10933. .11271
                /codon_start=1
                /transl_table=11
                /product="P450-like protein"
                /protein_id="BAB69191.1"
                /db_xref="GI:15823974"
                /translation="MEVGGQTIRAGEGVLCTLSTANRDEGAFTDPDELDLRRDARSHL
                SFGFGIHQCIGQSLARAEPQIALETLLRRLPGLRLAASFEELRFRDGVGFYGVKELPD
                PCPALPPDRP"
gene            11458. .29898
                /gene="olmA1"
CDS             11458. .29898
                /gene="olmA1"
                /codon_start=1
                /transl_table=11
                /product="modular polyketide synthase"
                /protein_id="BAB69192.1"
                /db_xref="GI:15823975"
                /translation="MACRLPGAAGPAEFLALLRNGEDAVGAAPGDRRYAPRRGGFLDS
                VDEFDAGFFGISPREAAAMDPQQRLVLELSWEALEDAGIPPARLGGSGAGVFVGAIAD
                DYATLSRAAGADAATPQTTTGLNRGIIANRVSYTLGLRGPSFTVDSGQSSSLVAVHLA
                AESLRRGECSVALAGGVNLILAEDSTVAVERFGALSPDGYCYTFDARANGYVRGEGGG
                VVVLKPLADAIADGDDIHCILAGSAVNNDGGGEGLTVPDRAGQEAVLTAAYEQAGISP
                AAVGYVELHGTGTPTGDPVEAAAVGAVLGAGRPAGQPLLVGSVKTNIGHLEGAAGIAG
                LLKAVLAVRHREIPATLHHASPNPAIPMAELGIGVNTALTPWPEQAGPLVVGVSSFGM
                GGTNCHVVLTEPAGYDAAARPATAGIRPAGTPAPLLVSGRDEQALRAQARKLRDHLDT
                HPGLPLRDVAHTLAAARAPLARRAVLLAADREDVSAGLARLADGTPGADLVRGTVSPG
                ALAFVFTGQGSQRPGMTAELYRTSAEYAAALDEVCASLDPWLRRPLRDVLFAAEGSDE
                AALLDRTEFTQPALFAVEVALFRYAEHCGLTPRFLLGHSVGELAAAHVAGVLSLADAC
                TLVAARGRLMQDRPATGAMVSVQATEDELAPFVSESVTLAALNGPASTVIAGDEDAVL
                AVAAHWAARGRKTKRLRVSHAFHSPHMDGMLTEFRRIAAGLTFGAPRIPIVSNETGTL
                LTEADARSPEYWVRHARAAVRFLDGVRLLEDQGVTTLLELGPDGVLSAMARDCLRDAG
                TVCVPLLRGRDRTEPEALLAALAGLYVRGVPMDWERTATEAGARRVPLPTYAFQRRRH
                WLPEPSASASRPVVNGGGHAGQAGHAGQAGRAGQAGHAGHDGEPAPAVRDRTTPELVR
                DRTTLELVRDATAQVLGHASPDDVATHEAFRELGFSSLMLAELGERLTEATGRRVPTT
                LLFDHPTPDALARRLEADGAERTDAGPVSPPAAAAHDDDPIVIVGMGCRLPGGVRSPE
                DLWRFVAAGGDAISALPVDRGWEFAADFPGEGGFLADVAEFDAGFFGISPREALAMDP
                QQRLLLETSWEALERAGVDALKLRGSRTGVFVGASPSEYGPRLHETSDADGHVLTGTA
                PSVLSGRIAYVLGLEGPALTVDTACSSSLVALHLAAQALRGGECDLALAGGVTVMATS
                GMFAEFARQGGLAPDGRCKAFADGADGTGWGEGVGMLAVQRLSDAVRDGRPVLAVVRG
                SAVNSDGASNGLTAPNGPSQQRVIRQALASAGLSAADVDMVEAHGTGTSLGDPIEAQA
                LLATYGQDRPDDRPLWLGSVKSNIGHTQAAAGVAGVIKSVLALRNGLLPKTLHVDEPT
                PEVDWSAGAVELLTEGREWPETDGPRRAGVSAFGISGTNAHVILEQAPSVEEPVTAPA
                DDGLLVPWVVSARSEEALRAQARRLADHVRESGLRPADVGLSLAVTRAGLEHRAVLVA
                EDEAGFLAHLDALAEGRTPAGVVVEGGTAFLFSGQGSQRAGMGRELYSAHPVFAAAFD
                EVCAAFDGLLPGALREVVFGGAGLDRTEWAQPALFALEVALFELVSSWGVRADVLLGH
                SIGELAAAYVAGVWSLADACRIVAARGRLMQALPEGGAMVAVEAAEDELPDLPGGVSV
                AAVNGPRSLVLSGDEEPVTALAQAFAEQGRRTKQLTVSHAFHSARMEPMLAEFAETLA
```

```
                AVEFRPPRIPVVSNVTGEMAGDEFTTPAYWVRHVREAVRFADGVGAAHAAGVRRYLEL
                GPDGTLTTLAHDTAHDAGRDVAAVPALRGDQPETRTVVTAVGRLYATGLTVDWPAFFA
                PYGARPTELPTYAFQHRRYWLEPVAPSSAPPANPLRYRTAWASVPDGPRPSLSGTWAV
                VVPASGSYEDIVAGVTAALRDHGARPVVVEVDPAGPERADVADTLRAALGEDAGGVVS
                LLGTAGSLSSGVAATAVLVQAWTDLGGGARLWCLTRGAVSVSPSDPLAAPAQAQLWGM
                GRVAALEHPERWGGLVDLPAVPDDRAWTRLCAVLAGGSGEDQVAVRAAGLFARRLGRA
                GTRADAAPAGPSPWRTDGTVLVTGGTGALGAHLAQWLAAAGAAHVLLTSRRGPDADGA
                AELTARLRETGTEVTVAACDVTDRDALAELLSGLPADRPLTGVVHAAGVLDDGVLDSL
                TPDRFAAVAGPKVTGAWHLHELTRDLDLSAFVLFSSFAGTVGLAGQANYAAANAYLDA
                LAVHRAQLGLPATAVAWGSWSGAGMAADTEAARRQLARTGLVPLAPDAALAALGQVMD
                DGETAVTVADVDWDRFAAGFGPGRPSPLLTGVPEVRRIEERETGRAGAPAETDRLGGL
```

```
  Query Match             45.4%;  Score 111.6;  DB 1;  Length 104326;
  Best Local Similarity   65.9%;  Pred. No. 3.9e-14;
  Matches  162;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              ||||||   | |||     || | |    ||| ||  ||| ||| || || || || ||
Db      38693 CTGGCCGCACTGACCGGACCCCAGCGCACGGCGGCCCTGCGCGACCTGGTGCGCGAACAG 38752

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              |  ||||| || ||| | ||||| ||| |     || |||  |||| ||   || ||
Db      38753 GCCGCGGCCGTACTCGGCCACGCCGACGCACGGTCCGTCGACCCCGCCCGCGCGCTCAAG 38812

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |  |||||||||| || ||||| || ||||  | || |||| ||||| ||  || ||
Db      38813 GAGCTCGGCTTCGACTCGCTCACCGCGGTCGAACTGCGCAACCGGCTGAGCACCGCGACC 38872

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || ||| | || || | | |  ||||||||||||||||||||    ||| |||| || |
Db      38873 GGCCTCAGGCTGCCCGCCACCATGGTCTTCGACCACCCGACCATCGCCGAACTGGCGGAT 38932

Qy        241 CACCTG 246
                ||||
Db      38933 TTCCTG 38938
```

```
RESULT 11
AB070949
LOCUS       AB070949               78210 bp    DNA     linear   BCT 11-OCT-2001
DEFINITION  Streptomyces avermitilis polyene macrolide biosynthetic gene
            cluster.
ACCESSION   AB070949
VERSION     AB070949.1  GI:15824136
KEYWORDS    .
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
  TITLE     Genome sequence of an industrial microorganism Streptomyces
            avermitilis: deducing the ability of producing secondary
            metabolites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
  MEDLINE   21477403
REFERENCE   2  (bases 1 to 78210)
  AUTHORS   Ikeda,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .78210
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
                     /note="Actinobacteridae"
     CDS             complement(1. .1254)
                     /codon_start=1
```

```
                   RSVVSASETAVTLVQQLSRLAETDRREVVLDLVAGQVAAVLGHAGTADIDPRRPLREL
                   GFDSLTAVELRNRLSAATGLRTAATVIFDHPTVDALAAHLLDGLMGSEAAAAADAPRT

  Query Match           45.6%;  Score 112.2;  DB 1;  Length 84985;
  Best Local Similarity   66.1%;  Pred. No. 3e-14;
  Matches  162;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||   | | |  |||| || |  || |  ||  |||||||  ||||||| | |||
Db      66476 CTTGCCGGGCTGTCCACGGCGGAGCGGCGCGAGGCGCTGCTCGATGTCGTACGGGCCCAG 66417

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
               | ||| | || ||  | ||||| | |||| || | ||||| || |||| | | ||||
Db      66416 ATCGCGACGGTCCTGGGCCACGCCGGCCCGGAAACGATCGCCCCTGACCGGGCCTTCCAG 66357

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |  |||||| |||| || || ||||| |||| |||| |||||||||    |||||| ||
Db      66356 GACCTCGGCCTCGACTCCCTGACGGCGATCGAACTCCGTAACCTGCTCGGCAAGGCCACC 66297

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || |   ||   || |||||| |||||||     |      || ||||
Db      66296 GGGCTGCGGCTCCCGGCAACGACCGTGTTCGACTACCCGACGGTGGATGCCCTGGCCGCC 66237

Qy        241 CACCT 245
              |||||
Db      66236 CACCT 66232


RESULT 10
AB070940
LOCUS       AB070940               104326 bp    DNA     linear   BCT 22-JAN-2002
DEFINITION  Streptomyces avermitilis oligomycin biosynthetic gene cluster.
ACCESSION   AB070940
VERSION     AB070940.1  GI:15823967
KEYWORDS    .
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
  TITLE     Genome sequence of an industrial microorganism Streptomyces
            avermitilis: deducing the ability of producing secondary
            metabolites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
  MEDLINE   21477403
REFERENCE   2  (bases 1 to 104326)
  AUTHORS   Ikeda,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .104326
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            complement(1. .1134)
                     /gene="mcmA"
     CDS             complement(1. .1134)
                     /gene="mcmA"
                     /note="methylmalonyl-CoA mutase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB69185.1"
                     /db_xref="GI:15823968"
                     /translation="MTGYSTAADANARFRQLVGSGTRAVPVVFDLPTRAGLDSDAPAA
                     RGLVGRRGVAVDSIDDMRVLFGGLRLDRVVPSLAVDAPAAPLLVLYQLVAEEHGAPAQ
                     RLAGAVRGDVFKEFLARGPGVFPLRPSLRLAMNVFAYCQAELPRWQCLSVCGHHLAGT
                     GADPALEVAFTVAGGIEYLRAAVTAGLDVRTVAPRLTLCLVGGTTGRENRAKLRATRR
                     VWTRTANERFGSGRTAPRLYVRTPRDGLATLACTEDIEAGATALVDRIERLGGTLTVL
                     EHGLAGQLLGRPHGDRPATPRPHPPHQDSLDLQARQTERLAKLRAWRVQPAVDEALLR
                     LRKAAQGDDNVLYPMREALAARATVGEVCDALRSTWGTGPSAT"
     gene            complement(1131. .1886)
                     /gene="pptA"
     CDS             complement(1131. .1886)
                     /gene="pptA"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphopantheteinyl transferase"
                     /protein_id="BAB69186.1"
                     /db_xref="GI:15823969"
                     /translation="MRAHEVFADVPVSSLHPREAALVQDAVESRRREFATGRHCARRA
                     LASLGVPPGPLLATPRGAPRWPAGVCGAITHCAGYRAAAVARAPGIRAVGIDAEPDRP
                     LPEGVLEAVALPAEQFQVKEHTARHPGVVWGKLLFSAKESVYKAWFPATGRPLGFEEA
                     LIDLDPVSSTFRVRVLPTALRAAEPELRDLHGRWLVRHGIIVTAVVVDGPGRDTPHAP
                     SPYDRTKHSAPTGRSADRPGGRPTDRATDRPTA"
     gene            complement(1910. .2875)
                     /gene="simX4 homolog"
     CDS             complement(1910. .2875)
                     /gene="simX4 homolog"
                     /codon_start=1
                     /transl_table=11
                     /product="SimX4 homolog"
                     /protein_id="BAB69187.1"
                     /db_xref="GI:15823970"
                     /translation="MSQDRHDGSTAALGGPSHGGRLLAVSDIHVRHRQNRDIVQDLHP
                     GCAEDWLIVAGDVAESVADITWALSLLRDRFAKVIWAPGNHELWTAPGDPDELRGERK
                     YRHLVDLCRALGVVTPEDPYPVWHGAGGPVAVAPLPLLYDYSFLPPGAWSKSHALRLA
                     DDAGVICTDEFLLHPDPYPSREAWCRARVRATRARLDALPDDLPTVLVNHFPLVREPT
                     RVLRHPEFALWCGTTATADWHLRYRAAAVVHGHLHIPRTIRCDDVPHHEVSLGYPREW
                     QRYGRRPDWPVQILTPAPAPAPAPVPVPVPAPPGHRTASPADRRP"
     gene            3213. .3971
                     /gene="olmC"
     CDS             3213. .3971
                     /gene="olmC"
                     /codon_start=1
                     /transl_table=11
                     /product="thioesterase"
                     /protein_id="BAB69188.1"
                     /db_xref="GI:15823971"
                     /translation="MSANRVDGNLWVRQFHPAPTAATKLVCLPHAGGSASYFLPVSKA
                     LAPGVDVLAIQYPGRQDRRHEKCIDNIAELADAVTAVLEPLADRPLVLFGHSMGATLA
                     FEVALRLERKGIVPLALFASGRRAPSRHRDESVHLRGDDEIIAELKTLSGTRSEVLGD
                     EEILRMVLPAIRSDYKAAETYRYAGGARLTAPIHAHFGVEDPRVSLDEAQAWADHTTG
                     HFELRSHPGGHFYLNDQAPRVIEEIARVSSSVAS"
     gene            4643. .7438
                     /gene="olmRI"
     CDS             4643. .7438
                     /gene="olmRI"
                     /codon_start=1
                     /transl_table=11
                     /product="luxR-family transcriptional regulatory protein"
                     /protein_id="BAB69189.1"
                     /db_xref="GI:15823972"
                     /translation="MLIGRGQEIDRLRAAYSSVRSTGRSRAVLVESAAGCGKSEILED
                     LVRHVESLGLIALRATAIAEERNTPLSVLRSLVNGAPFPDDLVRRFHTMSDAAARMIS
                     TEGRDKARRDAAWYELMRTFCTEVRALAQRAPLVVAVDDIHHGDADSLRQLLHLARHC
                     RNVPVLLTFTQSLAGGPADRTFGAELLRQPNLEHIRLPRFCRQDIADMLAEQGRADAG
                     QGTVERYFAVSGGNPLLLRALLEDARLEPDDTVDAQDVGPRAGQAYGRAVLACLDRGG
                     EKAHEVAAGLAVLGGAFTHELLSRLLDISAAALTENLRALRAAGIVVGTSFQHPAAKA
                     AVLDSLEPSHRLRLHQRSAALLHASGAPAALAAEHLHEAGQADGAWAVGVLQQAAEQA
                     LLADEARRAVSYLELAHAASADPRGRTEIKIRLGSVTRRVSAADAERHVDDALRGFGA
                     GLLTPSDTGGLAALLLAHGRLEQARSVLEQGRRPGDGDADRTAARRQLAEIRDSVLNR
                     GAWCAGTRAAWHENPAGRPSTGRAGTGLRAWLSGSHAALLAQRIAGTEGLLEITHLTD
                     ATFDLVLNAVWTLLALGAVERALHWCDTFLREAKRRDAPGWEAVFAAVRAEGALRGGN
                     LAEAERDARRCFAVLPERHRSALEGSAVSYLVTVQIARGKYEDAARQLSRPMADSLPD
                     TIHWLDYLRARGLYYMATRQYHSALRDFHEVGRLAERWGTDWPALLPWRTDTAEALLR
                     LGEREQGLRLATEQLAMVQDGAPRIKGISLRLQGLASEPAARLGLLNRAVVELSRADD
                     RLELSRTLFDLADAHRETRQTRQAAMVLRRARQLAKSCGAQPPRDRIGSPDAAGASGE
                     LFGEGPERNADWGHLSESERRVAGLAASGYTNREISVRLYITMSTVEQHLTRVYRKLK
                     INGREELLAEARAGVQV"
     gene            7658. .10486
```

```
gene            2190. .3566
                /gene="pimK"
CDS             2190. .3566
                /gene="pimK"
                /function="putative glucosyl transferase"
                /codon_start=1
                /transl_table=11
                /product="PimK protein"
                /protein_id="CAC20918.1"
                /db_xref="GI:12055069"
                /translation="MESARRPILFVSLPESGLLNPLLVLAGELSRQGVEDLWFATDEP
                RRNDVKRIAEGSPVEFASLGEVDSEMSAVTWSDEVYREVTQPSRFKAHRAVVRHTYRP
                GLQAEKFRRLQAVIDEVQPALMVIDCISGFAVDAAIARNIPYVLSVPFLPSNVLTAHT
                HFAKSYTPRGFPVPHTGLSRRMTLAQRVANELFKLRTFAMFLNPRLGKVLAEDNRRRN
                ELGLPKASFMARIEHADLVLCNSLAELDYPFDIPEKMRLVGAMVPPLPEAPDDQDLSR
                WLDAQSSVVYVGLGTITRLTREQVGSMVEVARRLEDRHQVLWKLPSEQQHLLPPRESL
                PGNLRVESWVPSQMDVLAHPHVKVFFTHGGGNGFNEGMYFGKPLVVRPLWVDCYDQAV
                RGQDFGLSLTLDRPQTIDVNDVVDKLTRVLGTPSFYEKAERRAALMRSAGGRETAAGL
                VLSLPALA"
gene            complement(3662. .9736)
                /gene="pimS4"
CDS             complement(3662. .9736)
                /gene="pimS4"
                /function="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="PimS4 protein"
                /protein_id="CAC20919.1"
                /db_xref="GI:12055070"
                /translation="MPNDKPNDKKLVDYLKWVTNDLHQTRQRLREVESAKQEPIAIVG
                MSCRLPGGVRSPEEFWQLLKDGRDGISTFPTDRGWDLDLLAGDGSGSSATAEGGFIDA
                ASFDAGFFGISPREAVTMDPQQRILLESSWEALERAGIDPVSLRGTETGVFVGTSGID
                YINTVMNSREDIEGHGSTGLIASILSGRVSYTFGLEGPAVTVDTACSSSLVSLHLAAQ
                ALRVGECSLALAGGVTAISTPMSFVGFSRQGGLAPDGRCKAFSDTADGTAWSEGVGVL
                VVERLSDARRNGHPVLAVVRGSAVNQDGASNGLTAPSGPSQQRVIRQALAAAGLSASD
                VDVVEAHGTGTALGDPIEAQALLATYGRDRDPERPLLLGSVKSNLGHTQAAAGAAGLI
                KMILAMRHGVLPRTLHVEAPSSHVDWSSGALQLLTEPQEWPKADRPWRAGVSSFAISG
                TNAHIILEQPPAAEEPEDEEGPHTSPSAVPWLVSAKSEEALTAQLDRISALTDRSPLD
                LGLSLATTRAHLEHRAVLLSEGGETVEVARGHAEHTAGKLAVLFSGQGAQRVGMGREL
                YGRFPVFAEALDGVLAHFDAGLRDVMFGEAEGLDETGFTQPALFAIEVALFRLSESLG
                VRPDFVAGHSIGEIAAAHVAGVFSLADACALVAARARLMQALPAGGAMVAVQAAESEV
                AARLVDGVSIAAVNGPEAVVIAGEEREVLRIAEEFTGLGRKTRRLPVSHAFHSPLMEP
                MLEDFRRVAEGLSYEAPRIALVSNLTGELATEDLVRSAEYWVRHVRETVRFADGVSAL
                AAEGASVFLELGPDGVLTAMAQHTLDGTATTVPALRKDRSEETALLTALAQLHVAGVG
                VDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVVGAGLRSAEHPLLGATLTLA
                ESGEVVLTGRLSLQAYPWLTDHTMGGSVLFPATGFLELAMRAADQVGCDRVDEFVLMT
                PLVLAEESGTQVQVLVGPADEAGTRTISIHARPDGDPDQPWMQHAGGTLTSGERIVDF
                GSTVWPPQGAEAVDLEGFYDTTGYGPAFQGLRSVWQCEDGAYVEVALPDQVAEDAGAF
                GLHPVLLDAVLQSRRMAGIGAAGDTFLPFAWRGVSLHAGGASVLRARVIKTGEDTLSL
                VAVDAEGAPVLSVEELVLRTQLTHTDAPLTARRGAERDALLSLEWVSGPEVPPAGVRC
                VSLGADELGVGASVATLADLSGDEDFVVVPVSGVGGDVPAAAHELTARTLGLLQEWSG
                REGTSASRLVFVTRNAVSADGGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADSG
                DLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGGT
                GGLGRELARHLVVERGVRHLLLVSRSGPEASGVGELRDELAAHGAEIAVRACDVADRA
                AVEAVVGAVDAAHPLTAVIHTAGVLDDGVIASLTPERLSAVLRPKVDAAWNLHEATKD
                LDLAAFVLYSSISGVLGSLGQGNYAAGNVFLDALAGHRAGLGLPAQSLAWGAWTQGGG
                MTATLSDADMQRMASYGTAPLSPERGLALFDAATAGSATHLVAVGRVSGPVRAQGPVP
                SMLRGLVKGARRTAVNTTGETAVSLADRLRAARPEERTQLMTDVVRTEAAAVLGHASS
                QNVDARREFYELGFDSLTSVELRNRLSTVTGLRLSATVVFDSKNPTDLAARLYADLAA
                QSALDGSASQAGFLPVAAPETDSLERMFLDALDSGKVPEAQRMLSALGALRPSFENTA
                ELEDLPLPATLAEGPGTPRLICVSTPTANGGVHEYARLAASFRGERHVSALPLVGFAT
                GERLPATAETAVRVIAESALRASDGNPFVLVGHSSAGAFAYLAAALLENTWGIRPEAV
                VLLDTLSLRHEQNESIDYAGLMRRHFMVDEVSPVRMTNSRLSAMARWMGMLNQLEVRH
                TTVPVLIIRATKETFGIGTDTGIYGEDHGSPVDVRSVDADHFSMVRDDAPETARIVRE
                WLDSLGNA"
gene            complement(9772. .15198)
                /gene="pimS3"
CDS             complement(9772. .15198)
                /gene="pimS3"
                /function="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="PimS3 protein"
                /protein_id="CAC20920.1"
                /db_xref="GI:12055071"
                /translation="MTTSSETVVAALRASLKEAENLRRQNRKLVAAATEPIAIVAMAC
                RYPGGVQSPEDLWELLTAGKDAIGAFPTDRGWDLDALRDAGVDRRGNAVSQEGGFLDG
                VGDFDADFFGISPRESVTMDPQQRLLLETSWEAIERAGIDPTTLRGSRTGVFIGTNGQ
                DYAYLLVRSLADATGDIGTGIAASATSGRLSYTLGLEGPAVTVDTACSSSLVALHTAS
                HALRAGECSLALVGGVNVMSSPGSLMEFSRQGGLAADGRCKAFSDTADGTGWSEGVGV
                LVVERLSDARRNGHPVLAVVRGSAVNQDGASNGFTAPSGPSQQRVIRQALAAAGLSAS
                DVDVVEAHGTGTPLGDPIEAQALLATYGQDRDPERPLLLGSVKSNLGHTQAAAGAAGL
                IKMILAMRHGVLPKTLYADVPSSHVDWTSGEVELLTEPQEWPKADRPWRAGVSSFGIS
                GTNAHVIVEQAEPDEEPVAEPGLAPAVVPWLVSAKTEEALAAQLARISSRRGTASPVD
                VGFSLASGRSAFEHRAVLLAQCPDGELSEAARGRAETSRPLAVLFSGQGAQRVGMGRE
                LYVRFPVFAEALDEVLAHFDAGLRDVMFGDAEGLDETGFTQPALFAIEVALFRLAESL
                GVRPDFVGGHSIGEIAAAHVAGVFSLTDACALVAARARLMQALPAGGAMVAVQASESE
                VAARLVDGVSIAAVNGPEAVVLAGDETAVLRIAEELAAGGRKTQRLPVSHAFHSPLME
                PMLEDFRRVAEGLSYGAPQIPLVSNVTGELAVEELVCSPDYWVRHVRETVRFADGVSA
                LEAEGASAFLELGPDGVLTAMARHTLDETKVAVPVLRKDRPEETALLTALAHLHVAGI
                GVDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVAGAGLRSAEHPLLGATLTL
                AESGEVVLTGRLSLQTHPWLTDHAIGGNVFFPGTGFLELAMHAADQAGCDRVEDLVLM
                TPLVLAEENGTQVQVLVGPADEAGTRTISIHARPEDDPDQPWMQHAGGTLTSGERVVD
                FGSTVWPPQDAVAVDLEGFYEGTGYGPAFQGLRSVWSCGDGAYAEVALPDHVAEDAGA
                FGLHPALLDAVLQSHRIAGVGSADDQFLPFAWRGVSLHAGGASVLRVRVVKTGDDSMS
                IAAVDAEGAPVLSAEAMVMRAYRDPGTGQAAARHGAERDELLSLEWVSGPEVPPAGVR
                CVSLGADELGVGASVASLADLSGDEDFVVLPVSGVGGDVPAAAHELTARTLGLLQEWS
                GREGSSGSRLVFVTRNAVSADEGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADS
                GDLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGG
                TGGLGRELARHLVVERGVRHLLLVSRSGPEAPGVSGLSDELAAHGAEVTVRACDVADR
                AAVHALVAAVPAEHPLTAVIHTAGVLDDGLVTSLTPERLSTVLRPKVDAAWNLHEATK
                DLDLAAFVLYSSISGVLGAPGQANYAAGNVFLDALAWHRMGQGLPAQSLAWGAWTQDG
                GMTATLSDTDMQRMASYGTAPLSRQRGLALFDAACARSAAHLVAVGRVGGPVRASAPV
                PSVLRGLVKGARRTAVNTTGETAVSLADRLRAARPEERPRLMTDLVRTEAAAVLGHAS
                AKNVDARREFNELGFDSLTSVELRNRLSALTGLRLSAALVFDYPNPTALADHFVSELV
                EEAATGPALLAELERLDSALAAGAPDARTRAAVTARLAQILDDWRDAPAEEDGAEVAE
                RLEAASTDEIFAFIDNELGRHSDR"
gene            complement(15210. .43733)
                /gene="pimS2"
CDS             complement(15210. .43733)
                /gene="pimS2"
                /function="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="PimS2 protein"
                /protein_id="CAC20921.1"
                /db_xref="GI:12055072"
                /translation="MDQEQKLRDYLKRASADLRRSRQRVGELEAAARERIAIVGMSCR
                FPGGVRSPEELWSLLAADGDGITKSPADRGWELLAEDAGEFSGGFLHDAAEFDASFFG
                ISPREALAMDPQQRLLLESSWEAFERAGIDPAAVRGTRTGMFVGAMPQEYRVGPDDDV
                QGFALTGTTTSVISGRLAYFFGAVGPAVTVDTACSSSLVALHLAAHSLRQGECSLALA
                AGVTVMSSPTTFVEFNRQGGLSADGRCRSFADSADGTGWSEGVGVLVLERLSEARRNG
                HEILAVIRGSAVNQDGASNGLTAPNGPSQRRVIEQALVSARLSADEVDVVEAHGTGTT
                LGDPVEAQALLATYGQQRDDDRPLLLGSVKSNLGHTQAAAGMAGVIKMVLAMRHGVLP
                RTLHVDAPSSHVDWTQGAVRLLTEHVPWPQGEQPRRAGVSSFGLSGTNAHTILEEAPA
                APSPEAGQEDTAPVAEPLAAGAVPWLVSGRTREALRAQAARLLEHLTSRPGLAPLDVA
                YSLATARSGLEHRAAFTAAGPDGARAALTALADGTPAPGLVQDTARTRSKLAFLFAGQ
                GSQRPGMGRELAARFPVFATALDEVLGHFDEGLERPLKDVLFAAEGTPEAALLDQTGY
                AQPALFAIEVALYRLAESFGITPDFLAGHSIGEIAAAHVAGVFTLADAAALVLARGRL
                MQALPEGGAMVSLEATEEEVLPLLEDRPDRLSIAAVNGPAAVVIAGEADDVTAVAEHF
                DALGRRTKRLRVSHAFHSPLMEPMLAEFRTAVSRLTPQAPVIPVVSGLTGTLATVEQL
                TSPDYWLDHARHAVRFADALRLARRARRRHLPRMGPDGVLSAMAQSCCRGHGHRRRRR
                LRHARRVRPGRPEAETLTAALAGLHTRGVAVQWEPYFQGTGARRADLPTYAFQRRRYW
                PKSLPATGGDVRAAGLGAAHHPLLTAAVSVANSDGLLLTGRLSRRTHPWLADHAVRGT
                VLLPGTAFLELAVRAGDEAGCGRVEELTLAAPLVLPEEGGVQVQVWVGSPDASGRRAV
                SVHSRPDGPEELPWTQHAAGTLVAGEHHAGFDATVWPPADATPLDLDGFYARMADTGF
                GYGPLFQGLRAAWRSGDDVYAEVALPDSDGAASTESSAAFGLHPALLDAALHAAAFVD
                LGEDARGGLPFSWQDVTLHASGASTVRVKLTPDGDDAVALAVADTTGAPVASIGSLVL
                RTLPDEQIDAAHSLVRDALFELRWSEARQTAHEAPASAAVLGPDLFGLERALAPTDLT
                VTAVPGPDAAPHAPDVVLAPVAGTTEDDVVTSVHARTAHVLERLRTWLTDERYAASRL
                VFVTRGALATDARQASDPVAAAVWGLVRAAQSEHPGRFGLLDLDPGTTEPATEPLLRA
                LALSSDEPQTAVRGTAVLTARLARAQAPQAATEWDPEGTVVLTGATGGLGRVLARHLV
                AERGVRHLLLVSRRGADADGAGELMAELAAHGAQVTLAACDLADRAAVRRLLSGVPAE
                HPVTAVVHSAGSLDDGVIASLTPERISAVLRPKADAVWHLHEATRDLDLAAFVVFSSL
                SGTVGAAGQGNYAAANAFLDALAQLRRAQGLPGLSLAWGPWAPTAGMTGGLTDDDLDR
                LARMGTPALTEEQGIALFDAATATDAAVLLPTRIDLSVLRVQSEIPPLWRGLIRTPAR
```

SPGRHARTHPRTGRGLFDAALAAGDAHLLPVRFDWASLRAQGEVPPLLRGLIRTRARR
SAVGGSAAAAGLVGRLSGRGTVERREVLLDLVRAQIAVVLGHANPETIESTRVFQDLG
FDSLTAVELRNRLNNATGLRLSATAVFDYPTADALVDFLLDELFGAQEEAELPAPVPS
PAGAADDPVVIVGMSCRYPGGVGSPEDLWRLVSEGVDAVSDFPTDRGWDVESLYSPDP
EALGTSYTRSGGFLHEAAEFDPDFFGMSPREALATDAQQRLLLETTWEAIERTGIDPA
SLRGSRTGVFAGVMYTDYGDLLVGDQFEGYRSNGSAASIASGRVSYTFGFEGPAVTVD
TACSSSLVALHWAAQSLRSGECSLAVAGGVTVMSTPTTFVEFSRQRGLSADGRCKAFA
DAADGVGWGEGVGMLVLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPAQQR
VIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLATYGQERPEDRPLLLGSVKS
NIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHVDWSAGAVELLTSEAEWPQG
EGPRRAGVSSFGVSGTNAHVILEQPGPDAADAAPDATVTDPGALAWVLSARNEAALRC
QAARLLSLVAGSDALCARDIGHSLVTGRSSFAHRAVVWGQDRDALVRALSALAVGEAD
AGLAEGASGAGRTAFLFSGQGSQRLGMGWELYARYPVFADAFDAVCAALDEHLERPLR
DVVWGEDAELLNQTAYAQAGLFAIEVALYRLAESWGMRPDFVAGHSIGEVAAAHVSGV
FSLPDACALVAARGRLMQQLPSGGAMMAIRATEDEVLPHLAEGVSLAAVNGPSSVVIS
GAEDAVLAIAAHFAGEGRKTTRLRVSHAFHSPLMEPMLEEFRAVVTRLSFGTPTIPVV
SNLTGRLAEPEQLAHADYWVRHVREAVRFADGIQALRAEGVTRFLELGPDGVLSAMAR
ESASDDAVLAPVLRRDRPEETALLGALAQLYVRGAHVDWTVPFAGSGARWADLPTYAF
QHERFWPSGGVARPGDVRSAGLGSAGHPLLGAAVELAGSGGLLFTGRLSVSSHPWLAD
HVVLGSVLVPGTALVELVLRAADEAGCDLLEELTLAAPLVLPASGAAVQVQVAVGEPD
EAGRRPVSVHAREGEGPWTLHASGAVTSGAEVPPFDATVWPPKGAEPVDVADCYDVLA
DAGLTYGPAFHGLQAAWKLGGDVYAEAKLPESTDGDAYGLHPALFDAALHASALGGAE
AGGVPFSWAGVSLHATGASHLRVRIREAGGALSVAIADTSGAPVASVESLVIRPLSAG
QVQAADRDALFKADWVPVPLTDERVEPGTGPEGEPLRTYADLDSLEGAAVPGTVLVAP
PSGAAGTVESVHAATVWALEMVQAWLADDRFATSRLVFVTRGAAFGADLAAAAVRGLV
RSAQSENPGRFGLVDMDGDADTTVPAQALATDEPELLVRGGEVLAARLVRAQSSHTVT
WDPSGTVLITGGTGGLGRSVARHLVSEHGVRSLLLVSRRGPAAEGAGELVAELRGSGA
EVVIEACDVTDAVAVADLVARHRISAVVHTAGVLDDGVVESLTPERLAVVLRPKVDAA
WNLHEATRGLDLDAFVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLA
WGPWSQDGGMTGTLSDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLA
ALRAQGEPQPLLRGLIRTRTRRSGAAAASGIAQRLAGLSTAERREALLDVVRAQIATV
LGHAGPETIAPDRAFQDLGLDSLTAIELRNLLGKATGLRLPATTVFDYPTVDALAAHL
LDELFGAETGTATETPLPVPGLPSLADDPVVIVGMSCRFPGGVASPEDLWRLVADGVD
AVSAFPTDRGWEIDDTYDPEREGAIATRSGGFLHDAAEFDPEFFGMSPREALTTDAQQ
RLLLETTWEALERAGMDPATLRGSRTGVFAGVMYHDYSTLLSGREFEGYQGSGSAGSV
ASGRVSYTFGFEGPAVTVDTACSSSLVALHLAAQSLRSGECSLALAGGVTVMSTPLTF
VEFSRQGGLSADGRCKAFADAADGVGWAEGAGILVLERLSDARRNGHRILATVRGSAV
NQDGASNGLTAPNGPAQQRVIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLA
TYGQERPEDRPLLLGSVKSNIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHV
DWSAGAVELLTSEAEWPQGEGPRRAGVSSFGISGTNAHVILEQPEPVAAETESITPDT
APDAAEDEAADSGTPVPALLSGRSASALRAQAARLLSRLDGDPGPRITDVAYSLATGR
SAFPHRAVILAANRADLLHSLSALAEGHTEAPAVVAQDRARSGKLAFLFSGQGSQRLG
MGRELYGRYPAFAEALDAVCAALDAHLDRPLRDVIWGEDAELLNRTGYAQTGLFAIEV
ALFRLLESWGVRPDHLLGHSIGEIAAAHVAGVLSLPDACALVAARGRLMQQLPSGGAM
MAIRATEDEVLPHLAEGVSLAAVNGPSSVVVSGAEDEVLALAAHFEEEGRKTTRLRVS
HAFHSPLMEPMLADFRAVADGMTYAAPRIPVVSNVTGRPATAEELCCAEYWVGHVREA
VRFADGVGALREQGVTTFLELGPDGSLSALAAESAADDSVLAPVLRKNRPEAPALLTA
LARLHAQGTPVDWSAAFAGTGARWVDLPTYAFQHERFWPSGGAARAGDVRSAGLGSAG
HPLLGAAVELAGSGGRLLTGRLSLSSHPWLADHVVLGSVLVPGTALMELVLRAADEVD
CAAVDELTLAAPLVLPASGAAIQVQVWVGEPDEAGRRPVSVHAREGEGPWTLHADGAL
APAAETVPFDTAIWPPQGAEHLDAAGCYERFADAGFAYGPVFQGLRAAWKLGEDIYAE
VALPEGTDGNAYGLHPALFDAALHAALLGGEGTDEAAVPFSWNGVTLHATGASRVRVR
IRPTEGGTSIALVDTAGAPVASVRSLTARPITAGQLQTGDRDSLFQVDWTTLHLTDER
ANSLALLGKDTEGILDTLSLQPHADLDDLAATGVHDTVLAPLPTRTAGTVESVHAATT
GALALIRSWLADDRFAASRLVFVTRGAVSGTDLAGASVWGLVRSALLEHPGRFGLVDV
DVDQDAEVPLVPRALASDEPQVLVRGGEVLAARLVRAQSSDTVTWDPSGTVLITGGTG
GLGRSVARHLVSEHGVRSLLLVSRRGPAAEGVDALVAELAECGAQVTVEACDVTDAVA
VADLVARHRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRGLDLDA
FVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLAWGPWSQDGGMTGTL
SDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLAALRAQGEPQPLLRG
LIRTPGRRTAAAATEGDTAAAFAGRLTGLSAAEGREVVLGAVRSQIAGVLGHAEATEI
DQDRAFLDLGFDSLTAVELRNRLGAVTGIRLPATLLFDYPTAELVAHLHARIAPEPT
VGPEALLGELERMEKSFGGLDLTEEMHEQIAGRLEVLRAKWDALRDTAAAAGHDGSPS
DEDFDFESASDDEVFDLLDNELGLS"

BASE COUNT 2611 a 7486 c 7397 g 2900 t
ORIGIN

Query Match 45.6%; Score 112.2; DB 6; Length 20394;
Best Local Similarity 66.1%; Pred. No. 4.1e-14;
Matches 162; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

```
Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||   | | |  |||| || |  || |  ||  |||||||  ||||||| | |||

Db      14677 CTTGCCGGGCTGTCCACGGCGGAGCGGCGCGAGGCGCTGCTCGATGTCGTACGGGCCCAG 14736

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
               | ||| | || ||  | ||||| | |||| || | ||||| || |||| | | ||||

Db      14737 ATCGCGACGGTCCTGGGCCACGCCGGCCCGGAAACGATCGCCCCTGACCGGGCCTTCCAG 14796

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              | |||||| |||| || || ||||| |||| |||| |||||||| |||||| ||

Db      14797 GACCTCGGCCTCGACTCCCTGACGGCGATCGAACTCCGTAACCTGCTCGGCAAGGCCACC 14856

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || |   ||   || |||||| |||||||    |    || ||||

Db      14857 GGGCTGCGGCTCCCGGCAACGACCGTGTTCGACTACCCGACGGTGGATGCCCTGGCCGCC 14916

Qy        241 CACCT 245
              |||||

Db      14917 CACCT 14921
```

RESULT 9
SNA278573/c
LOCUS SNA278573 84985 bp DNA linear BCT 17-MAY-2001
DEFINITION Streptomyces natalensis pimaricin biosynthetic gene cluster.
ACCESSION AJ278573
VERSION AJ278573.1 GI:12055067
KEYWORDS ABC transporter; cholesterol oxidase; cytochrome P450 monoxygenase; ferredoxin; glucosyl transferase; macrolide efflux pump; mycosamine dehyralase; orfX; pimA gene; pimB gene; pimC gene; pimD gene; pimE gene; pimF gene; pimG gene; pimH gene; pimI gene; pimJ gene; pimK gene; pimS0 gene; pimS1 gene; pimS2 gene; pimS3 gene; pimS4 gene; polyketide synthase; sensory transduction protein; thioesterase.
SOURCE Streptomyces natalensis.
ORGANISM Streptomyces natalensis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 84985)
AUTHORS Aparicio,J.F., Fouces,R., Mendes,M.V., Olivera,N. and Martin,J.F.
TITLE A complex multienzyme system encoded by five polyketide synthase genes is involved in the biosynthesis of the 26-membered polyene macrolide pimaricin in Streptomyces natalensis
JOURNAL Chem. Biol. 7 (11), 895-905 (2000)
MEDLINE 20547809
PUBMED 11094342
REFERENCE 2 (bases 1 to 84985)
AUTHORS Aparicio,J.F.
TITLE Direct Submission
JOURNAL Submitted (27-JUN-2000) Aparicio J.F., Molecular Genetics, Institute Of Biotechnology - Inbiotec, AV. REAL 1 - LEON, 24006, SPAIN
FEATURES Location/Qualifiers
source 1..84985
/organism="Streptomyces natalensis"
/db_xref="taxon:68242"
CDS <1..2086
/note="ORFX"
/codon_start=2
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC20917.1"
/db_xref="GI:12055068"
/translation="TRPDAFREATEPYGALLAPLMPERSADRVEGTKWTSDWSQARFL AHDAVCEVLLTLAGQHPLVLLLEDLQWADTASLDLLRLLSPRCQGHPLGIVLTARKFE TESDATLRRIVSDVLRGPRTATLRLCGLPRQAVGALVEAHAGSGVDPEVVEVLHRRSE GNPYFVMQLLSLVGDARRLHRPDAVDALLAHIPTGVREALNQRFAALPEPVLRVLRLC AVIGAEVDTDLLRRTATCDASVNAGLESAIRAGLLGEDPHHPGQLHFAHALVQETLFD ELPRADRRRLHAKVADALYARRRGHVGDEEIERVAHHSWHAKSALPAAEVLPRLLLAA EHAEQAMAYEQVEIWLRRAVHLVGFLPPDDPATVRLEQKLHIQLGQVLATTRGYGHPE AETALTRGRALSTVTHAPEDPSVLWALCAASLVTGRYDASRQLSGLLRDLSWQTREPV AVLGAAYGEGIVLHVRGMLPQALAELEHGVERADHFANEGHALARTFQHDPRVSCRSY NTFTHWLLGHRQIADARRRELLSLTEYESRPSDRSFALYVDGVVAAWEGDTRTALASS DEGVRVAGDHGLLYWKAMLGVVKGWGLVHAGDRDEGLALMHTSLAELGQSRTYLRHPL HLGLLGQAQHHNGQIEEARTTFRMLLAAVEQRRERVYLHPALPATPLLHELLGHGADE ALADRDPGRGQV"

```
                     EVVIEACDVTDAVAVADLVARHRISAVVHTAGVLDDGVVESLTPERLAVVLRPKVDAA
                     WNLHEATRGLDLDAFVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLA
                     WGPWSQDGGMTGTLSDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLA
                     ALRAQGEPQPLLRGLIRTRTRRSGAAAASGIAQRLAGLSTAERREALLDVVRAQIATV
                     LGHAGPETIAPDRAFQDLGLDSLTAIELRNLLGKATGLRLPATTVFDYPTVDALAAHL
                     LDELFGAETGTATETPLPVPGLPSLADDPVVIVGMSCRFPGGVASPEDLWRLVADGVD
                     AVSAFPTDRGWEIDDTYDPEREGAIATRSGGFLHDAAEFDPEFFGMSPREALTTDAQQ
                     RLLLETTWEALERAGMDPATLRGSRTGVFAGVMYHDYSTLLSGREFEGYQGSGSAGSV
                     ASGRVSYTFGFEGPAVTVDTACSSSLVALHLAAQSLRSGECSLALAGGVTVMSTPLTF
                     VEFSRQGGLSADGRCKAFADAADGVGWAEGAGILVLERLSDARRNGHRILATVRGSAV
                     NQDGASNGLTAPNGPAQQRVIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLA
                     TYGQERPEDRPLLLGSVKSNIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHV
                     DWSAGAVELLTSEAEWPQGEGPRRAGVSSFGISGTNAHVILEQPEPVAAETESITPDT
                     APDAAEDEAADSGTPVPALLSGRSASALRAQAARLLSRLDGDPGPRITDVAYSLATGR
                     SAFPHRAVILAANRADLLHSLSALAEGHTEAPAVVAQDRARSGKLAFLFSGQGSQRLG
                     MGRELYGRYPAFAEALDAVCAALDAHLDRPLRDVIWGEDAELLNRTGYAQTGLFAIEV
                     ALFRLLESWGVRPDHLLGHSIGEIAAAHVAGVLSLPDACALVAARGRLMQQLPSGGAM
                     MAIRATEDEVLPHLAEGVSLAAVNGPSSVVVSGAEDEVLALAAHFEEEGRKTTRLRVS
                     HAFHSPLMEPMLADFRAVADGMTYAAPRIPVVSNVTGRPATAEELCCAEYWVGHVREA
                     VRFADGVGALREQGVTTFLELGPDGSLSALAAESAADDSVLAPVLRKNRPEAPALLTA
                     LARLHAQGTPVDWSAAFAGTGARWVDLPTYAFQHERFWPSGGAARAGDVRSAGLGSAG
                     HPLLGAAVELAGSGGRLLTGRLSLSSHPWLADHVVLGSVLVPGTALMELVLRAADEVD
                     CAAVDELTLAAPLVLPASGAAIQVQVWVGEPDEAGRRPVSVHAREGEGPWTLHADGAL
                     APAAETVPFDTAIWPPQGAEHLDAAGCYERFADAGFAYGPVFQGLRAAWKLGEDIYAE
                     VALPEGTDGNAYGLHPALFDAALHAALLGGEGTDEAAVPFSWNGVTLHATGASRVRVR
                     IRPTEGGTSIALVDTAGAPVASVRSLTARPITAGQLQTGDRDSLFQVDWTTLHLTDER
                     ANSLALLGKDTEGILDTLSLQPHADLDDLAATGVHDTVLAPLPTRTAGTVESVHAATT
                     GALALIRSWLADDRFAASRLVFVTRGAVSGTDLAGASVWGLVRSALLEHPGRFGLVDV
                     DVDQDAEVPLVPRALASDEPQVLVRGGEVLAARLVRAQSSDTVTWDPSGTVLITGGTG
                     GLGRSVARHLVSEHGVRSLLLVSRRGPAAEGVDALVAELAECGAQVTVEACDVTDAVA
                     VADLVARHRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRGLDLDA
                     FVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLAWGPWSQDGGMTGTL
                     SDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLAALRAQGEPQPLLRG
                     LIRTPGRRTAAAATEGDTAAAFAGRLTGLSAAEGREVVLGAVRSQIAGVLGHAEATEI
                     DQDRAFLDLGFDSLTAVELRNRLGAVTGIRLPATLLFDYPTPAELVAHLHARIAPEPT
                     VGPEALLGELERMEKSFGGLDLTEEMHEQIAGRLEVLRAKWDALRDTAAAAGHDGSPS
                     DEDFDFESASDDEVFDLLDNELGLS"
BASE COUNT     2611 a   7486 c   7397 g   2900 t
ORIGIN

  Query Match           45.6%;  Score 112.2;  DB 1;  Length 20394;
  Best Local Similarity   66.1%;  Pred. No. 4.1e-14;
  Matches  162;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || |||   | | |  |||| || |  || |  ||  |||||||  ||||||| | |||
Db      14677 CTTGCCGGGCTGTCCACGGCGGAGCGGCGCGAGGCGCTGCTCGATGTCGTACGGGCCCAG 14736

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              | ||| | || ||  | ||||| | |||| || | ||||| || |||| | | ||||
Db      14737 ATCGCGACGGTCCTGGGCCACGCCGGCCCGGAAACGATCGCCCCTGACCGGGCCTTCCAG 14796

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |  |||||| |||| || || |||||  ||||  |||| |||||||||    |||||| ||
Db      14797 GACCTCGGCCTCGACTCCCTGACGGCGATCGAACTCCGTAACCTGCTCGGCAAGGCCACC 14856

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              || || || || || |   ||   || |||||| |||||||    |     || ||||
Db      14857 GGGCTGCGGCTCCCGGCAACGACCGTGTTCGACTACCCGACGGTGGATGCCCTGGCCGCC 14916

Qy        241 CACCT 245
              |||||
Db      14917 CACCT 14921


RESULT 8
AX067996
LOCUS       AX067996               20394 bp    DNA     linear   PAT 19-JAN-2001
DEFINITION  Sequence 3 from Patent WO0077222.
ACCESSION   AX067996
VERSION     AX067996.1  GI:12329806
KEYWORDS    .
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 20394)
  AUTHORS   Martin,J.F., Aparicio,J.F. and Colina,A.J.
  TITLE     Genes encoding enzymes in the biosynthesis of pimaricin and the
            application thereof
  JOURNAL   Patent: WO 0077222-A 3 21-DEC-2000;
            DSM N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .20394
                     /organism="Streptomyces natalensis"
                     /db_xref="taxon:68242"
     CDS             1. .20394
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC24642.1"
                     /db_xref="GI:12329807"
                     /translation="MSNEEKLREYLKRAIADLHETRQQLDETEAKQREPLAIVSMACR
                     FPGGVRSPEELWELLRDGVDAVSSFPRNRGWDLDALYHSDPAHQGTSYAREGGFLHDA
                     GEFDPGFFGISPREALAMDPQQRLLLETAWEAVERAGIDPESLAGSRTGVFVGTGHGG
                     YDAEGRRRADEVGGHLLTGNHISIASGRISYVLGLEGPALTVDTACSSSLVALHLAMH
                     ALRRDECAMALVGGATVMSTPQMFVEFSRQRGLAPDGRCKPFAAAADGTGWSEGVGLL
                     LVERLSDAVRNGYPVLAVLKGSAVNQDGASNGLTAPNGPSQQRVIRQALTGAGLAASD
                     IDAVEAHGTGTTLGDPVEAHALLATYGQQRAADRPCGLGSMKSNIGHTQAAAGIAGVM
                     KMVLAMRHGHLPRTLHLDEPTGHVDWSEGNARLLAEPEPWPSAGRPRRAAVSSFGISG
                     TNAHVILEQAPAHEAEPAPEPAARPGALPWILSARTEAGLRAQADRLGRHLRDRADLE
                     PAAVAHALADTRTLMEHRAVVVAGDREEFLRGLDALAAGRTANGLVSGVAVKAASAFL
                     FAGQGSQRPGMGRELHAAHPVFATAFDAVCAELDPHLDRPLRDIVFAEEGSAEAALLD
                     QTAYTQAALFALETALFRLVESWGVAPRFVAGHSIGELTAAHVSGVLTLHDAARLVAA
                     RGTLMQALPAGGAMVAVQATEDEIRERLAGHEDHVALAAANGPDSTVISGDEQAVTEI
                     AAHWEAQGRRTKRLRVSHAFHSPHMDDMLEDFRRVARGLTFHAPRIPVVSTVTGALAT
                     EDELRSPDYWVRQVRETVRFCAAVRTLEAEGVTTFVEIGTGGVLTPMVQDCLTTLEEP
                     VLVPLLRTGRPETVALTEGVATAFVHGVPVDRSAFPGAPGTSRADLPTYAFQRQWYWL
                     DPADHDEGEAAAAEAGEAGFWAAVEREDLQELSAVLAIDGSEADSLGSFLPTLSSWRR
                     QRRTQAAADRFSYRTHWAPRTASGGPTATGHWLVVLPEGGTDDPWTARLLDALNDQGL
                     HTDVRELPADHEPDAWGRHPVDGVLCLLALDERPTRSCPPYRRGLAATTNAAARPEGA
                     GIQAPLWCVTRGAVAVDRHEALKSPLQAQTWGLGRVAALESPQSWGGLIDLPDNLDGR
                     AVSALLSTLAGEEDQVAVRPAGVFARRLERITPGGDTGDRWSTHGTVLVTGGTGALGA
                     HLAHWLADAGAEHLVLTGRRGPQAPGAPELAAALTDRGVKVTLAACDAADRDALAAVL
                     ADIPPHLPLTGVVHAAGVLDDGVLDALTPERFETVLRPKARAAQNLHELTQDLDLDHF
                     VLFSSIVGVLGNAGQANYAAANAYLDALAEHRLAQGLPATSVSWGPGQAAAWHDSDAA
                     DRMSRDGLLPMAAAPRRRPAPALAQGMTQVTVADIDWSAYAPALTAVRPSPLIGDLPE
                     ARRALGPAEGPRRERSPLRDRIGALPPAEQEKAFLTMVREEAARVLGHPSPDTVDAQR
                     AFREQGFDSLMAVDLRNRLSAATGLRLPATLLFDHPTPLAAAACLRSEVLGAAGPATV
                     VQASTAALDEPVAIIGMACRFPGGVHSPEALWRLLAEGGDAITPMPADRGWDLDRLYH
                     PDPDHQGTSYARGGGFLDGAADFDADFFGISPREALAMDPQQRLLLETWEVLEQAGID
                     PESLRGSSTGVFAGTNTQDYGTALDAAQDEAGGHRLTGNAMSVVSGRVSYTFGFEGPA
                     LTVDTACSSSLVALHMAAQALRQGECSLAVAGGVTVMATPSSFVEFARQRGLAPDGRC
                     KPFAAAADGTGWSEGVGLLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLSAPSGP
                     SQQRVIRQALANARVAASEVDAVEAHGTGTTLGDPIEAQALLATYGQERPLLLGAVKS
                     NLGHTQAAAGVAGVMKMVLAMRHGMLPRTLHVDEPTGHVDWTAGAVELLTEHTDWPET
                     GHPRRAAVSAFGISGTNAHVVLELPAAEQPLVEQPSAAEPDAPATAPDRTPTASDGTA
                     PLLLSAKSESALRAQAARLHSHLERDPALRLTDAAYTLMTHRTAFAHRAAVRAADHEA
                     ALRALTALAAGEADPAVDTGTAHTGRDAVLFSGQGSQRIGMGRELSGRYPVFAEAFDT
                     VCAALDEHLDRPLRDVVRGEDEELLNRTVYAQAGLFAIEVALFRLVESWGVRPHYVAG
                     HSVGEIAAAHVAGVFSLADACALVAARGRLMQALPAGGAMAAIRATEDEVLPHLADSV
                     SIAAVNGPSSVVVSGAEHAVLSIAAHFEGAGRKTTRLRVSHAFHSPLMDPMLADFRAV
                     AEGLTYGEPELAVVSNVTGQLATPDQLRTPEYWVTHVRAAVRFADGIRALGAEGVTRF
                     LELGPDGVLSALARESAPDDAVCTPVLRKDRSEAATLLAALTHLHVHGTEIDWTAFLA
                     GRDAHAVDLPTYAFQHQRFWPTPDHTRTGDLGAVGLEATGHPLLSAAVELPDGEGLLF
                     TTRLSLQTHPWLAGHVVMGSVLLPGTAFAELALRAADEVGCDRVDELTLAAPLVLPEH
                     GGVQLQLRVGPADASGRRTLTARSRAEGDGDRPWVQHATGVLAEGESTPEPGYDFHTE
                     SWPPADAAPVELSGLYPDFAAHGFDYGPHFQGLRTAWRRGDEVFAEVALPAEAEGEAS
                     AYGLHPALLDAALHVVAFNGVDRGVVPFSWESVALHATGASAVRIRVVRHSGDTVSVD
                     VADTTGEPVASIGTLVLRAVSADQLAGGADPAVRDALFRVQWNPVRLPPAGAAVTVAT
                     LGSLAGAPFDGYPDLASLARSGRVAGAVLVPVEAGAGEVVADDVVGATHATAARALDL
                     ARSWLADDRFAASRLVFVTRGAVSGADLAGAAVWGLVRSALSEHPGRFGLVDLDDDAE
                     LALVPRVLASDEPQLLVRGGEVLAARLARAQSSHAVTWDPSGTVLVTGGTGGLGRVMA
                     RHLVVEHGVRNLLLVSRRGPAAEGAEELVTELRHSGAEVAVEACDVTDAAAVADLVAR
                     HRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRDLDLDAFVVFSSV
                     AGTIGSPGQANYAAGNAFLDALAHHRRAAGLPAASLAWGPWSRDGGMTGTLTDVDSSA
```

```
                    GTRADAAPAGPSPWRTDGTVLVTGGTGALGAHLAQWLAAAGAAHVLLTSRRGPDADGA
                    AELTARLRETGTEVTVAACDVTDRDALAELLSGLPADRPLTGVVHAAGVLDDGVLDSL
                    TPDRFAAVAGPKVTGAWHLHELTRDLDLSAFVLFSSFAGTVGLAGQANYAAANAYLDA
                    LAVHRAQLGLPATAVAWGSWSGAGMAADTEAARRQLARTGLVPLAPDAALAALGQVMD
                    DGETAVTVADVDWDRFAAGFGPGRPSPLLTGVPEVRRIEERETGRAGAPAETDRLGGL

  Query Match           46.1%;  Score 113.4;  DB 1;  Length 104326;
  Best Local Similarity  67.7%;  Pred. No. 1.6e-14;
  Matches  159;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              || ||  | | || | ||| ||| |  | ||   |  |||| |  |||||||| ||||
Db      83773 CTCGCGGAGCTGATGCCGGTCGAGCGGCAGGAGACCCTGCTGGCCCTCGTACGCGACCGG 83714

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
               || || |||||| |||||||   | || |    |||.| | | |||| | |||||||
Db      83713 ATCGCCGCCGTGCTCGGGCACGCCTCCTCGGACCGGATCGAGACGGACCGGCCGTTCCGC 83654

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |  |||||||||   ||||| || || || ||| | || |||| ||||| |  ||| ||
Db      83653 GACCTCGGCTTCACCTCACTGACCGCGGTGGAGCTGCGCAACCGGCTGAACGCGGCGACC 83594

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCG 235
              |||||||| || || ||    |||||||||||| | ||||| |  | | |||| |
Db      83593 GGACTCCGACTGCCCGTGAGCCTGGTCTTCGACTATCCGACGCTGGGCGAACTGG 83539


RESULT 7
SNA132222
LOCUS       SNA132222              20394 bp    DNA     linear   BCT 24-APR-1999
DEFINITION  Streptomyces natalensis pimS1 gene.
ACCESSION   AJ132222
VERSION     AJ132222.1  GI:4678702
KEYWORDS    pimaricin biosynthesis; pimS1 gene; polyketide synthase.
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 20394)
  AUTHORS   Aparicio,J.F., Colina,A.J., Ceballos,E. and Martin,J.F.
  TITLE     The biosynthetic gene cluster for the 26-membered ring polyene
            macrolide pimaricin. A new polyketide synthase organization encoded
            by two subclusters separated by functionalization genes
  JOURNAL   J. Biol. Chem. 274 (15), 10133-10139 (1999)
  MEDLINE   99214571
   PUBMED   10187796
REFERENCE   2  (bases 1 to 20394)
  AUTHORS   Aparicio,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-1999) Aparicio J.F., Microbiology, INBIOTEC, Av.
            Real, 1-Leon, 24006, SPAIN
FEATURES             Location/Qualifiers
     source          1. .20394
                     /organism="Streptomyces natalensis"
                     /db_xref="taxon:68242"
     gene            1. .20394
                     /gene="pimS1"
     CDS             1. .20394
                     /gene="pimS1"
                     /function="pimaricin biosynthesis"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="CAB41041.1"
                     /db_xref="GI:4678703"
                     /translation="MSNEEKLREYLKRAIADLHETRQQLDETEAKQREPLAIVSMACR
                     FPGGVRSPEELWELLRDGVDAVSSFPRNRGWDLDALYHSDPAHQGTSYAREGGFLHDA
                     GEFDPGFFGISPREALAMDPQQRLLLETAWEAVERAGIDPESLAGSRTGVFVGTGHGG
                     YDAEGRRRADEVGGHLLTGNHISIASGRISYVLGLEGPALTVDTACSSSLVALHLAMH
                     ALRRDECAMALVGGATVMSTPQMFVEFSRQRGLAPDGRCKPFAAAADGTGWSEGVGLL
                     LVERLSDAVRNGYPVLAVLKGSAVNQDGASNGLTAPNGPSQQRVIRQALTGAGLAASD
                     IDAVEAHGTGTTLGDPVEAHALLATYGQQRAADRPCGLGSMKSNIGHTQAAAGIAGVM
                     KMVLAMRHGHLPRTLHLDEPTGHVDWSEGNARLLAEPEPWPSAGRPRRAAVSSFGISG
                     TNAHVILEQAPAHEAEPAPEPAARPGALPWILSARTEAGLRAQADRLGRHLRDRADLE
                     PAAVAHALADTRTLMEHRAVVVAGDREEFLRGLDALAAGRTANGLVSGVAVKAASAFL
                     FAGQGSQRPGMGRELHAAHPVFATAFDAVCAELDPHLDRPLRDIVFAEEGSAEAALLD
                     QTAYTQAALFALETALFRLVESWGVAPRFVAGHSIGELTAAHVSGVLTLHDAARLVAA
                     RGTLMQALPAGGAMVAVQATEDEIRERLAGHEDHVALAAANGPDSTVISGDEQAVTEI
                     AAHWEAQGRRTKRLRVSHAFHSPHMDDMLEDFRRVARGLTFHAPRIPVVSTVTGALAT
                     EDELRSPDYWVRQVRETVRFCAAVRTLEAEGVTTFVEIGTGGVLTPMVQDCLTTLEEP
                     VLVPLLRTGRPETVALTEGVATAFVHGVPVDRSAFPGAPGTSRADLPTYAFQRQWYWL
                     DPADHDEGEAAAAAEAGEAGFWAAVEREDLQELSAVLAIDGSEADSLGSFLPTLSSWRR
                     QRRTQAAADRFSYRTHWAPRTASGGPTATGHWLVVLPEGGTDDPWTARLLDALNDQGL
                     HTDVRELPADHEPDAWGRHPVDGVLCLLALDERPTRSCPPYRRGLAATTNAAARPEGA
                     GIQAPLWCVTRGAVAVDRHEALKSPLQAQTWGLGRVAALESPQSWGGLIDLPDNLDGR
                     AVSALLSTLAGEEDQVAVRPAGVFARRLERITPGGDTGDRWSTHGTVLVTGGTGALGA
                     HLAHWLADAGAEHLVLTGRRGPQAPGAPELAAALTDRGVKVTLAACDAADRDALAAVL
                     ADIPPHLPLTGVVHAAGVLDDGVLDALTPERFETVLRPKARAAQNLHELTQDLDLDHF
                     VLFSSIVGVLGNAGQANYAAANAYLDALAEHRLAQGLPATSVSWGPGQAAAWHDSDAA
                     DRMSRDGLLPMAAAPRRRPAPALAQGMTQVTVADIDWSAYAPALTAVRPSPLIGDLPE
                     ARRALGPAEGPRRERSPLRDRIGALPPAEQEKAFLTMVREEAARVLGHPSPDTVDAQR
                     AFREQGFDSLMAVDLRNRLSAATGLRLPATLLFDHPTPLAAAACLRSEVLGAAGPATV
                     VQASTAALDEPVAIIGMACRFPGGVHSPEALWRLLAEGGDAITPMPADRGWDLDRLYH
                     PDPDHQGTSYARGGGFLDGAADFDADFFGISPREALAMDPQQRLLLETWEVLEQAGID
                     PESLRGSSTGVFAGTNTQDYGTALDAAQDEAGGHRLTGNAMSVVSGRVSYTFGFEGPA
                     LTVDTACSSSLVALHMAAQALRQGECSLAVAGGVTVMATPSSFVEFARQRGLAPDGRC
                     KPFAAAADGTGWSEGVGLLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLSAPSGP
                     SQQRVIRQALANARVAASEVDAVEAHGTGTTLGDPIEAQALLATYGQERPLLLGAVKS
                     NLGHTQAAAGVAGVMKMVLAMRHGMLPRTLHVDEPTGHVDWTAGAVELLTEHTDWPET
                     GHPRRAAVSAFGISGTNAHVVLELPAAEQPLVEQPSAAEPDAPATAPDRTPTASDGTA
                     PLLLSAKSESALRAQAARLHSHLERDPALRLTDAAYTLMTHRTAFAHRAAVRAADHEA
                     ALRALTALAAGEADPAVDTGTAHTGRDAVLFSGQGSQRIGMGRELSGRYPVFAEAFDT
                     VCAALDEHLDRPLRDVVRGEDEELLNRTVYAQAGLFAIEVALFRLVESWGVRPHYVAG
                     HSVGEIAAAHVAGVFSLADACALVAARGRLMQALPAGGAMAAIRATEDEVLPHLADSV
                     SIAAVNGPSSVVVSGAEHAVLSIAAHFEGAGRKTTRLRVSHAFHSPLMDPMLADFRAV
                     AEGLTYGEPELAVVSNVTGQLATPDQLRTPEYWVTHVRAAVRFADGIRALGAEGVTRF
                     LELGPDGVLSALARESAPDDAVCTPVLRKDRSEAATLLAALTHLHVHGTEIDWTAFLA
                     GRDAHAVDLPTYAFQHQRFWPTPDHTRTGDLGAVGLEATGHPLLSAAVELPDGEGLLF
                     TTRLSLQTHPWLAGHVVMGSVLLPGTAFAELALRAADEVGCDRVDELTLAAPLVLPEH
                     GGVQLQLRVGPADASGRRTLTARSRAEGDGDRPWVQHATGVLAEGESTPEPGYDFHTE
                     SWPPADAAPVELSGLYPDFAAHGFDYGPHFQGLRTAWRRGDEVFAEVALPAEAEGEAS
                     AYGLHPALLDAALHVVAFNGVDRGVVPFSWESVALHATGASAVRIRVVRHSGDTVSVD
                     VADTTGEPVASIGTLVLRAVSADQLAGGADPAVRDALFRVQWNPVRLPPAGAAVTVAT
                     LGSLAGAPFDGYPDLASLARSGRVAGAVLVPVEAGAGEVVADDVVGATHATAARALDL
                     ARSWLADDRFAASRLVFVTRGAVSGADLAGAAVWGLVRSALSEHPGRFGLVDLDDDAE
                     LALVPRVLASDEPQLLVRGGEVLAARLARAQSSHAVTWDPSGTVLVTGGTGGLGRVMA
                     RHLVVEHGVRNLLLVSRRGPAAEGAEELVTELRHSGAEVAVEACDVTDAAAVADLVAR
                     HRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRDLDLDAFVVFSSV
                     AGTIGSPGQANYAAGNAFLDALAHHRRAAGLPAASLAWGPWSRDGGMTGTLTDVDSSA
                     SPGRHARTHPRTGRGLFDAALAAGDAHLLPVRFDWASLRAQGEVPPLLRGLIRTRARR
                     SAVGGSAAAAGLVGRLSGRGTVERREVLLDLVRAQIAVVLGHANPETIESTRVFQDLG
                     FDSLTAVELRNRLNNATGLRLSATAVFDYPTADALVDFLLDELFGAQEEAELPAPVPS
                     PAGAADDPVVIVGMSCRYPGGVGSPEDLWRLVSEGVDAVSDFPTDRGWDVESLYSPDP
                     EALGTSYTRSGGFLHEAAEFDPDFFGMSPREALATDAQQRLLLETTWEAIERTGIDPA
                     SLRGSRTGVFAGVMYTDYGDLLVGDQFEGYRSNGSAASIASGRVSYTFGFEGPAVTVD
                     TACSSSLVALHWAAQSLRSGECSLAVAGGVTVMSTPTTFVEFSRQRGLSADGRCKAFA
                     DAADGVGWGEGVGMLVLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPAQQR
                     VIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLATYGQERPEDRPLLLGSVKS
                     NIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHVDWSAGAVELLTSEAEWPQG
                     EGPRRAGVSSFGVSGTNAHVILEQPGPDAADAAPDATVTDPGALAWVLSARNEAALRC
                     QAARLLSLVAGSDALCARDIGHSLVTGRSSFAHRAVVWGQDRDALVRALSALAVGEAD
                     AGLAEGASGAGRTAFLFSGQGSQRLGMGWELYARYPVFADAFDAVCAALDEHLERPLR
                     DVVWGEDAELLNQTAYAQAGLFAIEVALYRLAESWGMRPDFVAGHSIGEVAAAHVSGV
                     FSLPDACALVAARGRLMQQLPSGGAMMAIRATEDEVLPHLAEGVSLAAVNGPSSVVIS
                     GAEDAVLAIAAHFAGEGRKTTRLRVSHAFHSPLMEPMLEEFRAVVTRLSFGTPTIPVV
                     SNLTGRLAEPEQLAHADYWVRHVREAVRFADGIQALRAEGVTRFLELGPDGVLSAMAR
                     ESASDDAVLAPVLRRDRPEETALLGALAQLYVRGAHVDWTVPPAGSGARWADLPTYAF
                     QHERFWPSGGVARPGDVRSAGLGSAGHPLLGAAVELAGSGGLLFTGRLSVSSHPWLAD
                     HVVLGSVLVPGTALVELVLRAADEAGCDLLEELTLAAPLVLPASGAAVQVQVAVGEPD
                     EAGRRPVSVHAREGEGPWTLHASGAVTSGAEVPPFDATVWPPKGAEPVDVADCYDVLA
                     DAGLTYGPAFHGLQAAWKLGGDVYAEAKLPESTDGDAYGLHPALFDAALHASALGGAE
                     AGGVPFSWAGVSLHATGASHLRVRIREAGGALSVAIADTSGAPVASVESLVIRPLSAG
                     QVQAADRDALFKADWVPVPLTDERVEPGTGPEGEPLRTYADLDSLEGAAVPGTVLVAP
                     PSGAAGTVESVHAATVWALEMVQAWLADDRFATSRLVFVTRGAAFGADLAAAAVRGLV
                     RSAQSENPGRFGLVDMDGDADTTVPAQALATDEPELLVRGGEVLAARLVRAQSSHTVT
                     WDPSGTVLITGGTGGLGRSVARHLVSEHGVRSLLLVSRRGPAAEGAGELVAELRGSGA
```

```
                /gene="pptA"
                /codon_start=1
                /transl_table=11
                /product="phosphopantheteinyl transferase"
                /protein_id="BAB69186.1"
                /db_xref="GI:15823969"
                /translation="MRAHEVFADVPVSSLHPREAALVQDAVESRRREFATGRHCARRA
                LASLGVPPGPLLATPRGAPRWPAGVCGAITHCAGYRAAAVARAPGIRAVGIDAEPDRP
                LPEGVLEAVALPAEQFQVKEHTARHPGVVWGKLLFSAKESVYKAWFPATGRPLGFEEA
                LIDLDPVSSTFRVRVLPTALRAAEPELRDLHGRWLVRHGIIVTAVVVDGPGRDTPHAP
                SPYDRTKHSAPTGRSADRPGGRPTDRATDRPTA"
gene            complement(1910. .2875)
                /gene="simX4 homolog"
CDS             complement(1910. .2875)
                /gene="simX4 homolog"
                /codon_start=1
                /transl_table=11
                /product="SimX4 homolog"
                /protein_id="BAB69187.1"
                /db_xref="GI:15823970"
                /translation="MSQDRHDGSTAALGGPSHGGRLLAVSDIHVRHRQNRDIVQDLHP
                GCAEDWLIVAGDVAESVADITWALSLLRDRFAKVIWAPGNHELWTAPGDPDELRGERK
                YRHLVDLCRALGVVTPEDPYPVWHGAGGPVAVAPLFLLYDYSFLPPGAWSKSHALRLA
                DDAGVICTDEFLLHPDPYPSREAWCRARVRATRARLDALPDDLPTVLVNHFPLVREPT
                RVLRHPEFALWCGTTATADWHLRYRAAAVVHGHLHIPRTIRCDDVPHHEVSLGYPREW
                QRYGRRPDWPVQILTPAPAPAPAPVPVPVPAPPGHRTASPADRRP"
gene            3213. .3971
                /gene="olmC"
CDS             3213. .3971
                /gene="olmC"
                /codon_start=1
                /transl_table=11
                /product="thioesterase"
                /protein_id="BAB69188.1"
                /db_xref="GI:15823971"
                /translation="MSANRVDGNLWVRQFHPAPTAATKLVCLPHAGGSASYFLPVSKA
                LAPGVDVLAIQYPGRQDRRHEKCIDNIAELADAVTAVLEPLADRPLVLFGHSMGATLA
                FEVALRLERKGIVPLALFASGRRAPSRHRDESVHLRGDDEIIAELKTLSGTRSEVLGD
                EEILRMVLPAIRSDYKAAETYRYAGGARLTAPIHAHFGVEDPRVSLDEAQAWADHTTG
                HFELRSHPGGHFYLNDQAPRVIEEIARVSSSVAS"
gene            4643. .7438
                /gene="olmRI"
CDS             4643. .7438
                /gene="olmRI"
                /codon_start=1
                /transl_table=11
                /product="luxR-family transcriptional regulatory protein"
                /protein_id="BAB69189.1"
                /db_xref="GI:15823972"
                /translation="MLIGRGQEIDRLRAAYSSVRSTGRSRAVLVESAAGCGKSEILED
                LVRHVESLGLIALRATAIAEERNTPLSVLRSLVNGAPFPDDLVRRFHTMSDAAARMIS
                TEGRDKARRDAAWYELMRTFCTEVRALAQRAPLVVAVDDIHHGDADSLRQLLHLARHC
                RNVPVLLTFTQSLAGGPADRTFGAELLRQPNLEHIRLPRFCRQDIADMLAEQGRADAG
                QGTVERYFAVSGGNPLLLRALLEDARLEPDDTVDAQDVGPRAGQAYGRAVLACLDRGG
                EKAHEVAAGLAVLGGAFTHELLSRLLDISAAALTENLRALRAAGIVVGTSFQHPAAKA
                AVLDSLEPSHRLRLHQRSAALLHASGAPAALAAEHLHEAGQADGAWAVGVLQQAAEQA
                LLADEARRAVSYLELAHAASADPRGRTEIKIRLGSVTRRVSAADAERHVDDALRGFGA
                GLLTPSDTGGLAALLLAHGRLEQARSVLEQGRRPGDGDADRTAARRQLAEIRDSVLNR
                GAWCAGTRAAWHENPAGRPSTGRAGTGLRAWLSGSHAALLAQRIAGTEGLLEITHLTD
                ATFDLVLNAVWTLLALGAVERALHWCDTFLREAKRRDAPGWEAVFAAVRAEGALRGGN
                LAEAERDARRCFAVLPERHRSALEGSAVSYLVTVQIARGKYEDAARQLSRPMADSLPD
                TIHWLDYLRARGLYYMATRQYHSALRDFHEVGRLAERWGTDWPALLPWRTDTAEALLR
                LGEREQGLRLATEQLAMVQDGAPRIKGISLRLQGLASEPAARLGLLNRAVVELSRADD
                RLELSRTLFDLADAHRETRQTRQAAMVLRRARQLAKSCGAQPPRDRIGSPDAAGASGE
                LFGEGPERNADWGHLSESERRVAGLAASGYTNREISVRLYITMSTVEQHLTRVYRKLK
                INGREELLAEARAGVQV"
gene            7658. .10486
                /gene="olmRII"
CDS             7658. .10486
                /gene="olmRII"
                /codon_start=1
                /transl_table=11
                /product="luxR-family transcriptional regulatory protein"
                /protein_id="BAB69190.1"
                /db_xref="GI:15823973"
                /translation="MLASAGQGEGGVVLVNAGPGVGRTALLDAFLRKATESGARVCAA
                AGSSAETEAEFGVVGQLFPPTGPLGAGFRLARAIGSRDAGPGLDADRLFDALCQELRR
                RPTVIAVDDAQLADAASLRFLRYLVRRLRTAPVMMVLTEPTGPDALPPSFQAELLRHP
                QCRRLGLLPLSRDGVTQVVSAQVGSDRVPRLAADFHAISGGNPLLVHALLADHRAAHR
                LSPVPSVPTVPSVSSASSAASASSAASAASAVSAAPAASAAFAKAALAWAYRDEPVLF
                DVACGFAVLGESATPARVACLVDRGADAVDQVMTALDTAGLLAGGTFRSALVGKALLE
                HVDVETRAELRRRAARLLHGDGAPATVVARHLLAAPVAEPWAGRTLLAAAEKAMQEGR
                EELSLDCLRLAGQSATTEQERAAVVMARVKIGWEIDPRLVSPWLDELVAALRRGDLRS
                GDAAWIVKHLAWQDRFPEAVESLSTLAAAGDAGEPGVEAELHFVRHWLRHSCPPLEAS
                VPGPSNPTRPRTPRLRRVSPASYAGEVLGKVLTEGPDDHTVAMAEEILRGCRYSETVV
                EALEAALLALVYAGRPDRAAFWCDGLLQQAGTRSTGTSVAILSGIRAEIALRQGALPA
                VEQYARSALSAISLPGWGVAVGSPLAALVHSTTASGMSGVAGAWLNQAVPDGMFRTRH
                GLLYTHARGHYHLAVDRPAAALDDFLTCGGLAKEWGMDVPTFLPWRTSAAQAQLALGN
                PGQARALVREQLGRPGGSTARARGVALRLLAATGEPGRRPALLRESVNLLESCGDQVE
                LLRALADRCQALDHVGAPAKARMTARHARNVADACGGDALFRKLFRDEVPDREGDPDG
                LTGEEQGSASLTAAERRVTALAAALGHSNREIGRKLFITKSTVEQHLTRVYRKLGVRNR
                ADLGDLLAGTNLGAGTPAMAMASSAVAG"
CDS             10933. .11271
                /codon_start=1
                /transl_table=11
                /product="P450-like protein"
                /protein_id="BAB69191.1"
                /db_xref="GI:15823974"
                /translation="MEVGGQTIRAGEGVLCTLSTANRDEGAFTDPDELDLRRDARSHL
                SFGFGIHQCIGQSLARAEPQIALETLLRRLPGLRLAASFEELRFRDGVGFYGVKELPD
                PCPALPPDRP"
gene            11458. .29898
                /gene="olmA1"
CDS             11458. .29898
                /gene="olmA1"
                /codon_start=1
                /transl_table=11
                /product="modular polyketide synthase"
                /protein_id="BAB69192.1"
                /db_xref="GI:15823975"
                /translation="MACRLPGAAGPAEFLALLRNGEDAVGAAPGDRRYAPRRGGFLDS
                VDEFDAGFFGISPREAAAMDPQQRLVLELSWEALEDAGIPPARLGGSGAGVFVGAIAD
                DYATLSRAAGADAATPQTTTGLNRGIIANRVSYTLGLRGPSFTVDSGQSSSLVAVHLA
                AESLRRGECSVALAGGVNLILAEDSTVAVERFGALSPDGYCYTFDARANGYVRGEGGG
                VVVLKPLADAIADGDDIHCILAGSAVNNDGGGEGLTVPDRAGQEAVLTAAYEQAGISP
                AAVGYVELHGTGTPTGDPVEAAAVGAVLGAGRPAGQPLLVGSVKTNIGHLEGAAGIAG
                LLKAVLAVRHREIPATLHHASPNPAIPMAELGIGVNTALTPWPEQAGPLVVGVSSFGM
                GGTNCHVVLTEPAGYDAAARPATAGIRPAGTPAPLLVSGRDEQALRAQARKLRDHLDT
                HPGLPLRDVAHTLAAARAPLARRAVLLAADREDVSAGLARLADGTPGADLVRGTVSPG
                ALAFVFTGQGSQRPGMTAELYRTSAEYAAALDEVCASLDPWLRRPLRDVLFAAEGSDE
                AALLDRTEFTQPALFAVEVALFRYAEHCGLTPRFLLGHSVGELAAAHVAGVLSLADAC
                TLVAARGRLMQDRPATGAMVSVQATEDELAPFVSESVTLAALNGPASTVIAGDEDAVL
                AVAAHWAARGRKTKRLRVSHAFHSPHMDGMLTEFRRIAAGLTFGAPRIPIVSNETGTL
                LTEADARSPEYWVRHARAAVRFLDGVRLLEDQGVTTLLELGPDGVLSAMARDCLRDAG
                TVCVPLLRGRDRTEPEALLAALAGLYVRGVPMDWERTATEAGARRVPLPTYAFQRRRH
                WLPEPSASASRPVVNGGGHAGQAGHAGQAGRAGQAGHAGHDGEPAPAVRDRTTPELVR
                DRTTLELVRDATAQVLGHASPDDVATHEAFRELGFSSLMLAELGERLTEATGRRVPTT
                LLFDHPTPDALARRLEADGAERTDAGPVSPPAAAAHDDDPIVIVGMGCRLPGGVRSPE
                DLWRFVAAGGDAISALPVDRGWEFAADFPGEGGFLADVAEFDAGFFGISPREALAMDP
                QQRLLLETSWEALERAGVDALKLRGSRTGVFVGASPSEYGPRLHETSDADGHVLTGTA
                PSVLSGRIAYVLGLEGPALTVDTACSSSLVALHLAAQALRGGECDLALAGGVTVMATS
                GMFAEFARQGGLAPDGRCKAFADGADGTGWGEGVGMLAVQRLSDAVRDGRPVLAVVRG
                SAVNSDGASNGLTAPNGPSQQRVIRQALASAGLSAADVDMVEAHGTGTSLGDPIEAQA
                LLATYGQDRPDDRPLWLGSVKSNIGHTQAAAGVAGVIKSVLALRNGLLPKTLHVDEPT
                PEVDWSAGAVELLTEGREWPETDGPRRAGVSAFGISGTNAHVILEQAPSVEEPVTAPA
                DDGLLVPWVVSARSEEALRAQARRLADHVRESGLRPADVGLSLAVTRAGLEHRAVLVA
                EDEAGFLAHLDALAEGRTPAGVVVEGGTAFLFSGQGSQRAGMGRELYSAHPVFAAAFD
                EVCAAFDGLLPGALREVVFGGAGLDRTEWAQPALFALEVALFELVSSWGVRADVLLGH
                SIGELAAAYVAGVWSLADACRIVAARGRLMQALPEGGAMVAVEAAEDELPDLPGGVSV
                AAVNGPRSLVLSGDEEPVTALAQAFAEQGRRTKQLTVSHAPHSARMEPMLAEFAETLA
                AVEFRPPRIPVVSNVTGEMAGDEFTTPAYWVRHVREAVRFADGVGAAHAAGVRRYLEL
                GPDGTLTTLAHDTAHDAGRDVAAVPALRGDQPETRTVVTAVGRLYATGLTVDWPAFFA
                PYGARPTELPTYAFQHRRYWLEPVAPSSAPPANPLRYRTAWASVPDGPRPSLSGTWAV
                VVPASGSYEDIVAGVTAALRDHGARPVVVEVDPAGPERADVADTLRAALGEDAGGVVS
                LLGTAGSLSSGVAATAVLVQAWTDLGGGARLWCLTRGAVSVSPSDPLAAPAQAQLWGM
                GRVAALEHPERWGGLVDLPAVPDDRAWTRLCAVLAGGSGEDQVAVRAAGLFARRLGRA
```

```
                     entries PF00109 ketoacyl-synt, Beta-ketoacyl synthase,
                     PF00698 Acyl_transf, Acyl transferase domain and PF00550
                     pp-binding, Phosphopantetheine attachment site and matches
                     to Prosite entries PS00606 Beta-ketoacyl synthases active
                     site, PS00340 Growth factor and cytokines receptors family
                     signature 2 and PS00012 Phosphopantetheine attachment
                     site"
                     /codon_start=1
                     /transl_table=11
                     /product="putative type I polyketide synthase"
                     /protein_id="CAC22144.1"
                     /db_xref="GI:12274801"
                     /db_xref="SPTREMBL:Q9EX54"
                     /translation="MSNEQKMVEYLKWVTTDLQKARQRLAELEAGNDEPVAIVGMACR
                     FPGGVASPDDLWRLVMDEADGITDWPTDRGWDIDAIYDPERGKPGKSYTRKGGFLDGA
                     AQFDPGFFGISPREALAMDPQQRVLLETAWETFEQAGIDPATLRGSQVGVFAGVVEQS
                     YLGLEGPEELEGYLMTSKLSSVASGRIAYTFGFEGPAVSVDTACSSSLVALHLAVQSV
                     RKGESTMALAGGSTVSGNPSGFVDFSRQNGLAEDGRIKSFAAAADGTAWSEGVGLLLV
                     EKLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQERVIHQALAGARLNASDVD
                     VVEAHGTGTRLGDPIEAQALLASYGRQRPADRPLYLGSLKSNIGHTVAAAGAGGVIKM
                     IMAMRHGILPKTLNVDEPTPMVDWEAGAVELLTEARPWPETGAPRRAGVSAFGVSGTN
                     AHVILEEPPAEEPEEPADTAVTALPVLPWALSARTAPALRDQARRLLSHVEERPAEDP
                     LDVAYSLAAERSALEHRAVVVGSDREELLTGLRALAEGRTTPTGAVQDTRQTGKTAFL
                     FTGQGAQRTGMGMDLYRTFPAYAHAFDTITAQLDPHLDQPLHHTITTGHHLHHTGNTQ
                     PALFATEVALYRLLETWGITPDYLAGHSIGELTAAHISGILTLQDACTLVTARARLMQ
                     NLPTHGTMIALQATEQEILPHLTGHEHHLTIAAINSPTSLVISGNQTAAQHIATQLQA
                     QGRKTKTLTVSHAFHSPHMDGMLHDFHHTATQLTYHEPTIPIVSTLTGNLATHNDLRT
                     PTYWTDQLRNTVRYTQAIHTLHTAGVTTYTEIGPDATLTPLTQQTLDTTTAIPLLRHN
                     QPETHTLLTAISRLHNRGIPINWKNLFTTTPAKHTPLPTYAFQHERFWVETSGVAADA
                     SDLGLESAGHPMLGAAVTVAGADQVLFTSRLSLRTHPWLADHVVLDQTLVPATALVEM
                     AIRAGDELGSTVLNDLSVLGPLVLPGKGGVHVQVSVGSPDASGLRELSIYSRPDNADV
                     PWTLNARGRLSFRGPDAPFSLAQWPPAGAEEVPLDEAYDKLAVQGYVYGPVYRGLEHV
                     WRLGDEVYADVRLPDGSRTGIGGFGLHPALLETALAASCLAGYGSSPEGAVWISTDWK
                     DVRLHATGADRLRVRLAPDADGALNVQLADRAGRPVASVSSVACRAIAVGEVSNALAR
                     NRDSLFHVDWAPFPVAPSAGTERWVSLGADEPGARRYPDVASLARAVAAGESVDVALV
                     GTETGDTGDRVADVHDVTREALALVQDWLAEERLAGTRLVVRTRGGVTTAGDEDVTDL
                     GAAAVWGLLRSAQSENPGRIVLVDVDDEEASARILPAVVASGEPQSAVRDGQVFVPRL
                     GRVPEAPAGAATGAWDRRGTVLITGGTGSLGGLFARHLVVEHGVRNLLLTSRRGRDAE
                     GAAELEAELTALGARVEIAACDAADRAALAAVLADIPAEYPLTGVLHTAGVLDDGLIG
                     SQTPQRLSAVLRPKVDATWNLHELTRDMDLSAFVLFSSIAAVVGGPGQSTYSAANHFL
                     DALAEHRRSRGLAATSLAWGLWSQTSGMTGHLDEADLKRIARSGFRPVESEQGPALLD
                     IALTVGRAGLVATPMDLTALREQVQVPVLLSKLARTPVRRTAQNAEVSSESLAERLAG
                     LGDAEQREAVLGIVLGQAAVVLGHSDATGIDRQRSFTALGFDSLTSVELRNQLGTSTG
                     LRLPPTLVFDHPTPVALAEYLHGEVRAELSADGSASRPGVDFAAEVRLADDVRPADEV
                     ITTAADPKEILLTGASGFLGAFLLRDLMRTTTARVHCLVRGADEAAAMERLKANAEWY
                     RVWDDIDPERVSVVLGDLAEPRLGLDEESFDALARTVDVVYHNGARVHWLLPYETLKA
                     SNVTGTEEVLRLAARHRTVPVHYVSTVGVFDGVREPGVPLRVTDPTGPAESLPSGYLR
                     SKWVAEQVIEVARDRGLPVSVYRVDVISGDRVNGACQTRDFVWLTLKGLIQARSVPKG
                     TEGRFHLLPVDYVSAAITGISRQPGTVGRTFHLFNQSSLALSQCVELLRSLGYELDEV
                     DWDTWTQVVTSGDNALLPLLDAFEMMTSDTDGFYPPIDTAETVAALEGTGIGIPVLTR
                     ELFEKYVAFFVEEGHFPAAG"
     misc_feature    complement(4732. .4935)
                     /gene="SC2C4.03c"
                     /note="Pfam match to entry PF00550 pp-binding,
                     Phosphopantetheine attachment site, score 68.50, E-value
                     2.5e-17"
     misc_feature    complement(4810. .4857)
                     /gene="SC2C4.03c"
                     /note="PS00012 Phosphopantetheine attachment site"
     misc_feature    complement(5248. .5268)
                     /gene="SC2C4.03c"
                     /note="PS00340 Growth factor and cytokines receptors
                     family signature 2"
     misc_feature    complement(7342. .8286)
                     /gene="SC2C4.03c"
                     /note="Pfam match to entry PF00698 Acyl_transf, Acyl
                     transferase domain, score 248.50, E-value 9.5e-71"
     misc_feature    complement(8596. .9867)
                     /gene="SC2C4.03c"

Query Match           50.2%;  Score 123.4;  DB 1;  Length 24225;
Best Local Similarity  69.0%;  Pred. No. 1.8e-16;
Matches 169;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              ||||| | | | | ||||||| |     |   ||| |    || ||||| || | |||
Db      15698 CTGGCTCCGCTGTCCGCGGCCGAGCGCGAGCGGGCACTCACGGATCTCGTGCGCGTCCAG 15639

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              || ||||| |||||| ||||| | |||||    || |||| | ||| || | | ||||
Db      15638 GTCGCGGCGGTGCTCGGGCACTCTGACCCCGGCGCGATCGAGTCCGGCCGGGCCTTCCAG 15579

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              |  || |||||||| ||||| || |||||||  | || |||| ||||| ||  || |
Db      15578 GAGCTGGGCTTCGACTCACTGACAGCCGTCGAACTTCGCAACCAGCTGAGCACCGCGAGC 15519

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              ||||| ||||| ||   | | || |||||||||||||||  |||| |||   |||||||
Db      15518 GGACTGCGCCTGCCCACCACCCTCGTCTTCGACCACCCCTCCCCCGCCGCTCTCGCCGCC 15459

Qy        241 CACCT 245
              |||||
Db      15458 CACCT 15454


RESULT 6
AB070940/c
LOCUS       AB070940              104326 bp    DNA     linear   BCT 22-JAN-2002
DEFINITION  Streptomyces avermitilis oligomycin biosynthetic gene cluster.
ACCESSION   AB070940
VERSION     AB070940.1  GI:15823967
KEYWORDS    .
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
  TITLE     Genome sequence of an industrial microorganism Streptomyces
            avermitilis: deducing the ability of producing secondary
            metabolites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
  MEDLINE   21477403
REFERENCE   2  (bases 1 to 104326)
  AUTHORS   Ikeda,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .104326
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            complement(1. .1134)
                     /gene="mcmA"
     CDS             complement(1. .1134)
                     /gene="mcmA"
                     /note="methylmalonyl-CoA mutase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB69185.1"
                     /db_xref="GI:15823968"
                     /translation="MTGYSTAADANARFRQLVGSGTRAVPVVFDLPTRAGLDSDAPAA
                     RGLVGRRGVAVDSIDDMRVLFGGLRLDRVVPSLAVDAPAAPLLVLYQLVAEEHGAPAQ
                     RLAGAVRGDVFKEFLARGPGVFPLRPSLRLAMNVFAYCQAELPRWQCLSVCGHHLAGT
                     GADPALEVAFTVAGGIEYLRAAVTAGLDVRTVAPRLTLCLVGGTTGRENRAKLRATRR
                     VWTRTANERFGSGRTAPRLYVRTPRDGLATLACTEDIEAGATALVDRIERLGGTLTVL
                     EHGLAGQLLGRPHGDRPATPRPHPPHQDSLDLQARQTERLAKLRAWRVQPAVDEALLR
                     LRKAAQGDDNVLYPMREALAARATVGEVCDALRSTWGTGPSAT"
     gene            complement(1131. .1886)
                     /gene="pptA"
     CDS             complement(1131. .1886)
```

```
Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6155 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 6214

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6215 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 6274

Qy        241 CACCTG 246
              ||||||
Db       6275 CACCTG 6280
```

RESULT 5
SC2C4/c
LOCUS SC2C4 24225 bp DNA linear BCT 12-MAY-2002
DEFINITION Streptomyces coelicolor cosmid 2C4.
ACCESSION AL512902 AL645882
VERSION AL512902.2 GI:20520683
KEYWORDS accA1; secreted FAD-binding protein; type I polyketide synthase.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2)
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 24225)
AUTHORS Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J., Kinashi,H. and Hopwood,D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
PUBMED 8843436
REFERENCE 2 (bases 1 to 24225)
AUTHORS Seeger,K.J. and Harris,D.
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 24225)
AUTHORS Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
TITLE Direct Submission
JOURNAL Submitted (15-JAN-2001) Streptomyces coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT On May 9, 2002 this sequence version replaced gi:12274798.
Notes:
Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web.
(URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for CDS which show significant similiarity to other CDS in the database.
The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nih.go.jp/
jun/cgi-bin/frameplot.pl. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 2C4.
FEATURES Location/Qualifiers
source 1. .24225
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 2C4"
misc_feature complement(2. .1477)
/note="previously sequenced DNA fragment. EMBL:AF113603 Streptomyces coelicolor putative acyl-CoA carboxylase complex A subunit (accA1) gene, complete cds."
misc_feature 2. .119
/note="nominal overlap with Streptomyces coelicolor cosmid SCAH10"
misc_feature complement(374. .1471)
/note="Pfam match to entry PF00289 CPSase_L_chain, Carbamoyl-phosphate synthase (CPSase), score 585.30, E-value 3.8e-172"
misc_feature complement(602. .625)
/note="PS00867 Carbamoyl-phosphate synthase subdomain signature 2"
RBS complement(1485. .1489)
RBS 1774. .1778
gene 1786. .3438
/gene="SC2C4.02"
/note="SCO6272"
CDS 1786. .3438
/gene="SC2C4.02"
/note="SC2C4.02, possible secreted FAD-binding protein, len: 550 aa; similar to TR:Q9ZAR8 (EMBL:U80222) Streptomyces peucetius ceasius DnrW, 485 aa; fasta scores: opt: 1323 z-score: 1315.4 E(): 0; 45.3% identity in 470 aa overlap. Contains Pfam match to entry PF01565 FAD_binding_4, FAD binding domain. Also contains possible N-terminal region signal peptide sequence"
/codon_start=1
/transl_table=11
/product="putative secreted FAD-binding protein"
/protein_id="CAC22143.1"
/db_xref="GI:12274800"
/db_xref="SPTREMBL:Q9EX55"
/translation="MTEVSRRKLMKGAAVSGGALALPALGAPPATAAPAAGPEDLPGP
AAAAAGRRPAPAPVHVTRSDPRYQGLISGSNQRWVGNPDYIRVVSSAGQVVRAVQEAV
DKGLKIAVRSGGHCDEDFVANRDVRVVIDMAGMDSVTYDRERRAFAVGPGARLGTVYR
TLYKRWGVVLPGGTCPTVGAGGHITGGGYGALSRSRGLTVDHLYAVEVVHVDARGRAR
KVVATREEDDPNRELWWAHTGAGGGNFGVITRYWLRSPDAVGDDPSTLLPAAPSEVLL
SDVSWSWDDLDEASFTRLLRNFTEWHARNSAPDSPGRFLFSQLKTMHKAAGYFRMSTQ
VDAAAPDADRLLDDYLAAISEGTGVTYHVGDRYRAPWLYAVTEWSGFVEASVPRWKSK
SAYVREVMPEEQLRAVYRQLTRDDYPGPYGMIAIVGFGGKINEVAPGDTATAQRDSIA
KMLYCSLWSDPADDALHQRWIREAYEDVYASTGGVPRPGGVNDGCYINYADADLADPA
LNRSGIPWHELYFKGNYPRLQRVKATWDPRNVFSHRLGIRKP"
misc_feature 1948. .2535
/gene="SC2C4.02"
/note="Pfam match to entry PF01565 FAD_binding_4, FAD binding domain, score 6.50, E-value 1.1e-06"
gene complement(3514. .9972)
/gene="SC2C4.03c"
/note="SCO6273"
CDS complement(3514. .9972)
/gene="SC2C4.03c"
/note="SC2C4.03c, probable type I polyketide synthase, len: 2152 aa; N-terminal domain almost identical to TR:Q9L8J0 (EMBL:AF202898) Streptomyces coelicolor A3(2) type I polyketide synthase (fragment), 720 aa; fasta scores: opt: 4356 z-score: 4494.6 E(): 0; 93.8% identity in 721 aa overlap and C-terminal domain similar to SW:HETM_ANASP (EMBL:L22883) Anabaena sp. polyketide synthase HetM, 506 aa; blastp scores: Score= 761 (267.9 bits), Expect= 2.2e-74, Sum P(2)= 2.2e-74; Identities= 184/491 (37%), Positives= 278/491 (56%). Also highly similar, in this same cosmid, to the N-terminal region of SC2C4.04c and to SC2C4.05c. Contains Pfam matches to

EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGAL
ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
TAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVVIADIDWSKIEHTSQTSDLVSA
ARELREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
MTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
NLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY
VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGG
VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
LRCPDAGAGGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPE
DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
VSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
SLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLGLS"

gene 12071. .30790
/gene="aveA2"

CDS 12071. .30790
/gene="aveA2"
/note="contains Ave module 3, module 4, module 5 and module 6
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 2"
/protein_id="BAA84475.1"
/db_xref="GI:5902892"
/translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGINP
EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
ACQALRSGECTMALAGGASVMATPFVFTEFSRQRGLAADGRCKAFSAAADGTGWSEGV
GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
NDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALHRL
LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
TQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
NRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLTAVPADEPIAIV
GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
LTYHDYAARFPTAPAGFEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
HLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSAAADGTGWGE
GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAD
LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAADDVPGEAAGDDAGTGGEATGPA
AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH
GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
QGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
PQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
QHDQARINTWTYQETWKPLTPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG
TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
GTWQGNGLATGQVSEHLRRRGMFAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA
AVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ
FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI
GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAAL
LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
TLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
KATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD
IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL

Query Match 100.0%; Score 246; DB 1; Length 64957;
Best Local Similarity 100.0%; Pred. No. 4e-42;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
Db 6035 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 6094

Qy 61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
Db 6095 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 6154

```
Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6812 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 6871

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6872 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 6931

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6932 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 6991

Qy        241 CACCTG 246
              ||||||
Db       6992 CACCTG 6997
```

```
RESULT 3
E38020
LOCUS       E38020                30690 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Avermectin aglycon synthase gene.
ACCESSION   E38020
VERSION     E38020.1  GI:18626909
KEYWORDS    JP 2000245457-A/1.
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 30690)
  AUTHORS   Omura,S. and Ikeda,H.
  TITLE     Avermectin aglycon synthase gene
  JOURNAL   Patent: JP 2000245457-A 1 12-SEP-2000;
            THE KITASATO INSTITUTE
COMMENT     OS   Streptomyces avermitilis
            PN   JP 2000245457-A/1
            PD   12-SEP-2000
            PF   24-FEB-1999 JP 1999046961
            PR
            PI   SATOSHI OMURA,HARUO IKEDA
            PC   C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC   (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(11916)
            FT   CDS             (11971). .(30687).
FEATURES              Location/Qualifiers
     source           1. .30690
                      /organism="Streptomyces avermitilis"
                      /db_xref="taxon:33903"
BASE COUNT     5355 a  12455 c   8617 g   4263 t
ORIGIN
```

```
  Query Match           100.0%;  Score 246;  DB 6;  Length 30690;
  Best Local Similarity  100.0%;  Pred. No. 4.7e-42;
  Matches  246;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5935 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 5994

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5995 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 6054

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6055 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 6114

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6115 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 6174

Qy        241 CACCTG 246
              ||||||
Db       6175 CACCTG 6180
```

```
RESULT 4
AB032367
LOCUS       AB032367              64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
            Sciences, Kitasato University, Microbial chemistry; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES              Location/Qualifiers
     source           1. .64957
                      /organism="Streptomyces avermitilis"
                      /db_xref="taxon:33903"
     gene             101. .12019
                      /gene="aveA1"
     CDS              101. .12019
                      /gene="aveA1"
                      /note="contains a loading module, Ave module 1 and module
                      2
                      multifunctional polyketide synthase"
                      /codon_start=1
                      /transl_table=11
                      /product="type I polyketide synthase AVES 1"
                      /protein_id="BAA84474.1"
                      /db_xref="GI:5902891"
                      /translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
                      ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
                      AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
                      VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTR
                      MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
                      MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
                      RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
                      DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
                      LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
                      GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
                      MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
                      GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
                      GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
                      RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
                      GDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
                      LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
                      AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAG
                      LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
                      EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
                      VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
                      AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
                      AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
                      ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
                      PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
                      THHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
```

```
        and polyketide synthase gene from Streptomyces avermitilis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11096)
  AUTHORS   Hong,Y.-S. and Lee,J.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2000) Anticancer Agent Research Laboratory, Korea
            Research Institute of Bioscience and Biotechnology, P.O. Box 116,
            YuSong-Gu, Taejon 305-600, South Korea
FEATURES             Location/Qualifiers
     source          1. .11096
                     /organism="Streptomyces avermitilis"
                     /strain="ATCC31271"
                     /db_xref="ATCC:31271"
                     /db_xref="taxon:33903"
     CDS             220. .>11096
                     /note="Avr1"
                     /codon_start=1
                     /transl_table=11
                     /product="avermectin polyketide synthase"
                     /protein_id="AAG09812.1"
                     /db_xref="GI:9964076"
                     /translation="MADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAE
                     FAPYVDWSVEQVLRDSPDAPGLDRVDVVQRTLFAVMISLAALWRSQGVEPCAVLGHSL
                     GEIAAAHVSGGLSLADAARVGDAWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDN
                     PARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEER
                     VVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLQQGP
                     KTFVEMSPHPVLTMGLQELAADLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGH
                     GETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHL
                     GFDSVMGVELRNRLSKATGLRLPVTLILDHTTPAAVAARLRTAALGHLDEDTAPVPDS
                     PSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQR
                     HAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERA
                     GIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRSAYS
                     FGLEGPAITVDTGCSASLVTLHLACQSLRSGECTFALAGGVSDVHPGMFIEFSRQCGL
                     SVDGRCKAVSAAADGTGWGEGIGLLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQAL
                     ANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMA
                     AAGVGGVIKMVMALREGVLPRTLHVDEPSPQGLAAGAVRLLTEAVPWPGDAAGRLRRA
                     GVSSFGIGGTNAHVILEEAPAAGGCVAGGRVLEGAPGLAISVAESVAAPVAVSAPVAE
                     SVPVPVPVPVPVPVSARSEAGLRAQAEALRQYEAVQPDVSLADVGAGLACRQAVLEHH
                     VVILAACTSSSRSAARTTARSSSTARPQARPAPTSARRIGALAAGSGSAALTTGHAPG
                     GDRGGVVFVFPWQGGQWAGMGVRLLCLLRVFARRMQACEEALAPWVDWSVVDILRRDA
                     GDAVWEQADVVQPVLFSVMVSLAALWRSYGIEPNEVLGHSKDEIAAAHIYGALSLKDA
                     AKTVALPPQEVEQLIGERGGRLWVAAVNGPRSTAVSGDAEAVVEVLAYCAGTGVRPRI
                     PVDYASHCPHVQPLREELLELLGDISPQPYGVPFFSTVEGTWLDTTTLDAAYWYRNLH
                     QPVRFSHDVQALADQGHRVLLEVSPHPTLVPAIEDTTEDTARRRHCDRQPPPRRERHP
                     LLPQRLRLDPYYRHRQTHHVAPLLDPPRHLPPPLDAPRPAHLPFQHQHYWLESSQPGA
                     GSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVV
                     PALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNIL
                     TNLHHHGITPIPLPLTTHHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPH
                     PHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLAR
                     TTLLEHPTHTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTL
                     TPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLT
                     TQLQQKGIHLTITTCDTSTPRPTHNNSLNTIPPQHPVTTVIHTGGILDDATLTNLTPT
                     QLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAH
                     HRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIADTER
                     PYVVIADIDWSKIEHTSQTSDLVSAARERERPAVQRPTPPAELHKTLAHQTSADQRAAL
                     LELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFD
                     HPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDF
                     WDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGIS
                     PREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQS
                     TDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMAL
                     AGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRN
                     GHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGT
                     TLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLL
                     PRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEA
                     PADDVPGGPPAGEGDAGSDDEAAAGSPGVPWPWLVSAKSQPALRAQAQALHAHLTDHPG
                     LDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGT
                     GEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQN
                     DNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITGAHLAG
                     ILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSL
                     VISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHT
                     PLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHH
                     NLPNTPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHAQPTPHPHPLDLPTYP
                     FQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLA
                     DHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAAD
                     GKGRRLVSVYARGGSACGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAV
                     AVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPAL
                     LNGVVQPLSVLLPGGTGFGKGAGFGKGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGG
                     GREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVP
                     VCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGGVGEVVGGVLGVVQG
                     WLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDT
                     DTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLTRLESSGDVPAQRS
                     GDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLA
                     GVCGVRDLLLVSRRGPDAPGAEGLRRSWPRGAEVRIVACDVGERREVVRLAGGCSCRV
                     SVFVDFSAVASVTVRLPVASDVRKEAAMAYATVEEFTDYLDPDP"
BASE COUNT     1756 a    4059 c    3528 g    1753 t
ORIGIN

  Query Match             100.0%;  Score 246;  DB 1;  Length 11096;
  Best Local Similarity   100.0%;  Pred. No. 5.8e-42;
  Matches  246;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6055 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 6114

Qy         61 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6115 GTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAGTCGTTCCGT 6174

Qy        121 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6175 GCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATCAAGGCAACA 6234

Qy        181 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6235 GGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAACTCGCCGTA 6294

Qy        241 CACCTG 246
              ||||||
Db       6295 CACCTG 6300


RESULT 2
AX006889
LOCUS       AX006889               12381 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 1 from Patent WO0001827.
ACCESSION   AX006889
VERSION     AX006889.1  GI:9994904
KEYWORDS    .
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 12381)
  AUTHORS   Kellenberger,J.L., Leadlay,P.F., Staunton,J., Mcarthur,H.A. and
            Stutzman-Engwall,K.J.
  TITLE     Polyketides, their preparation, and materials for use therein
  JOURNAL   Patent: WO 0001827-A 1 13-JAN-2000;
            KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
            STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PFIZER (US); MCARTHUR
            HAMISH ALASTAIR IRVIN (US); STUTZMAN ENGWALL KIM JONELLE (US)
FEATURES             Location/Qualifiers
     source          1. .12381
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     1884 a    4561 c    4005 g    1931 t
ORIGIN

  Query Match             100.0%;  Score 246;  DB 6;  Length 12381;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-42;
  Matches  246;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6752 CTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTACGAGACCAT 6811
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 477.19 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_5935_6180
Perfect score: 246
Sequence: 1 ctggcccatcagacgtcggc.....ccaaactcgccgtacacctg 246

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

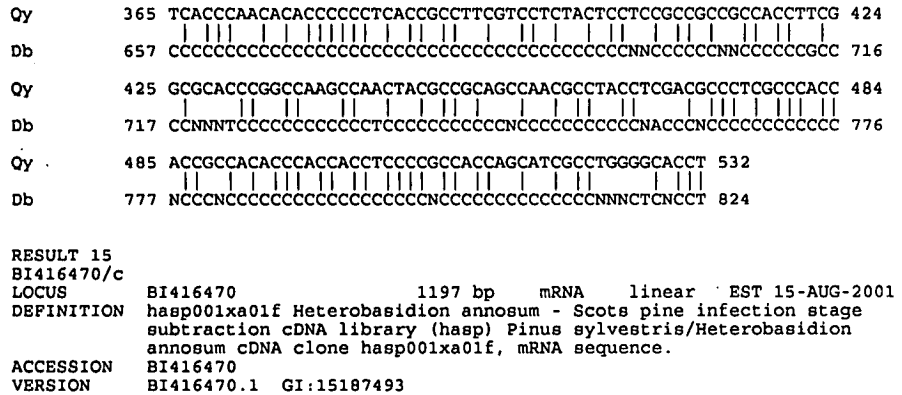
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		246	100.0	11096	1	AF275943	AF275943 Streptomy
2		246	100.0	12381	6	AX006889	AX006889 Sequence
3		246	100.0	30690	6	E38020	E38020 Avermectin
4		246	100.0	64957	1	AB032367	AB032367 Streptomy
5	c	123.4	50.2	24225	1	SC2C4	AL512902 Streptomy
6	c	113.4	46.1	104326	1	AB070940	AB070940 Streptomy
7		112.2	45.6	20394	1	SNA132222	AJ132222 Streptomy
8		112.2	45.6	20394	6	AX067996	AX067996 Sequence
9	c	112.2	45.6	84985	1	SNA278573	AJ278573 Streptomy
10		111.6	45.4	104326	1	AB070940	AB070940 Streptomy
11		110.8	45.0	78210	1	AB070949	AB070949 Streptomy
12		110.2	44.8	11480	1	STMPKS3ORF	L09654 Streptomyce
13		110.2	44.8	24568	1	AF220951	AF220951 Streptomy
14		110.2	44.8	50937	6	AR159871	AR159871 Sequence
15		110	44.7	123580	1	AF263912	AF263912 Streptomy
16		110	44.7	125401	6	AX211739	AX211739 Sequence
17		107.4	43.7	31422	6	E38021	E38021 Avermectin
18	c	107.4	43.7	64957	1	AB032367	AB032367 Streptomy
19		106.2	43.2	113193	1	AF357202	AF357202 Streptomy
20		105.2	42.8	65140	6	AX211705	AX211705 Sequence
21		103.6	42.1	1681	6	AR173228	AR173228 Sequence
22		103.6	42.1	37948	1	AF079138	AF079138 Streptomy
23	c	102	41.5	39314	1	SGR300302	AJ300302 Streptomy
24		102	41.5	43280	1	SFU78289	U78289 Streptomyce
25		102	41.5	53784	1	AMM223012	AJ223012 Amycolato
26		102	41.5	53789	6	A69720	A69720 Sequence 3
27		102	41.5	90445	1	AF040570	AF040570 Amycolato
28		101	41.1	30000	6	AX250263	AX250263 Sequence
29		101	41.1	32870	1	AF007101	AF007101 Streptomy
30		101	41.1	77534	1	AF235504	AF235504 Streptomy
31		100.4	40.8	2700	6	AX211733	AX211733 Sequence
32		100.4	40.8	2706	1	AF420594	AF420594 Streptomy
33		100.4	40.8	9513	6	AX089460	AX089460 Sequence
34		100.4	40.8	41782	1	AY118081	AY118081 Streptomy
35		100.4	40.8	50000	6	AX089417	AX089417 Sequence
36		100.4	40.8	50000	6	AX089420	AX089420 Sequence
37		100.4	40.8	80161	1	AY007564	AY007564 Saccharop
38		100.4	40.8	80161	6	AR165018	AR165018 Sequence
39		99.4	40.4	30000	6	AX250262	AX250262 Sequence
40		98.8	40.2	5676	6	A69718	A69718 Sequence 1
41		98.8	40.2	41097	1	AF016585	AF016585 Streptomy
42	c	97.8	39.8	1434	6	AR173224	AR173224 Sequence
43	c	97.6	39.7	77534	1	AF235504	AF235504 Streptomy
44	c	97.2	39.5	30000	6	AX250263	AX250263 Sequence
45		96.8	39.3	6459	6	AX089458	AX089458 Sequence

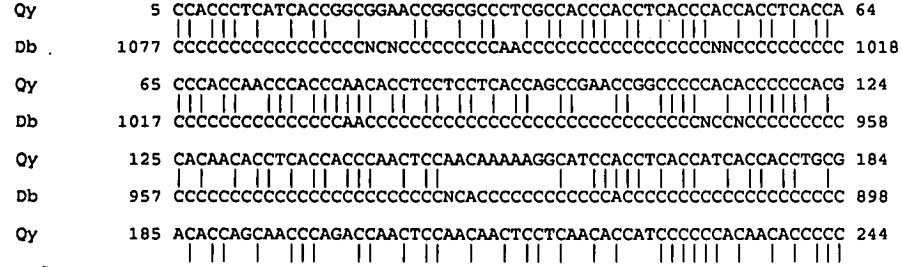
ALIGNMENTS

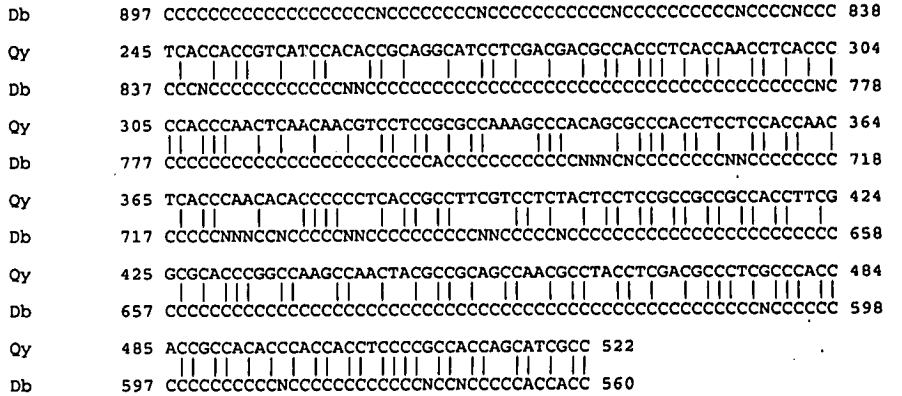
RESULT 1
AF275943
LOCUS AF275943 11096 bp DNA linear BCT 02-SEP-2000
DEFINITION Streptomyces avermitilis avermectin polyketide synthase gene, partial cds.
ACCESSION AF275943
VERSION AF275943.1 GI:9964075
KEYWORDS .
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 11096)
AUTHORS Hong,Y.-S. and Lee,J.J.
TITLE Targeted Gene Disruption of the avermectin O-methyltransferase gene



RESULT 15
BI416470/c
LOCUS BI416470 1197 bp mRNA linear EST 15-AUG-2001
DEFINITION hasp001xa01f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xa01f, mRNA sequence.
ACCESSION BI416470
VERSION BI416470.1 GI:15187493
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 1197)
AUTHORS Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026,S-750 07, Uppsala, Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES Location/Qualifiers
source 1. .1197
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="hasp001xa01f"
/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
/note="Vector: pT-Adv; Site_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidon annosum (FP5)."
BASE COUNT 20 a 28 c 1036 g 28 t 85 others
ORIGIN

Query Match 29.0%; Score 155; DB 13; Length 1197;
Best Local Similarity 53.7%; Pred. No. 1.1e-16;
Matches 278; Conservative 0; Mismatches 240; Indels 0; Gaps 0;





Search completed: June 18, 2003, 01:00:17
Job time : 687.626 secs

/strain="C57BL/6J"
/db_xref="niaEST:H3060F02-3"
/db_xref="taxon:10090"
/clone="H3060F02"
/clone_lib="NIA Mouse 15K cDNA Clone Set"
/sex="Clones arrayed from a variety of cDNA libraries"
/dev_stage="Clones arrayed from a variety of cDNA libraries"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearrayed set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse develolpment, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

BASE COUNT 157 a 858 c 103 g 108 t 2 others
ORIGIN

Query Match 29.1%; Score 155.4; DB 12; Length 1228;
Best Local Similarity 56.3%; Pred. No. 9.7e-17;
Matches 291; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

```
Qy          6 CACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCAC 65
              | ||| |  | |  || | |   ||| | | || |||| | | | ||||  |   || |
Db        660 CCCCCCCCCCCCACGCCGCCCTCCCGCGCCCCCCCCCCCAGCCCTCGCACCCGCCGCCCCC 719

Qy         66 CCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACGC 125
              || ||  ||| |||| |||| || || ||||  |   ||  || || | || |||| |
Db        720 CCCCCCGCCCCCCCAGCACCCCCCCCCCCACCACGCACCCCCCCCCACCCCCCCGCCCACCC 779

Qy        126 ACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCGA 185
              || | ||    | ||   | ||| |    |   | | ||  | ||   ||| ||  |
Db        780 CCACCCCCCGCACGCCACCCCCCACCCCCACTCCCCCAACGCCCCCCCAACCCCCGCCCC 839

Qy        186 CACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCCT 245
              ||||  |  ||||  ||  | |   |  | || |  | |   |||||     |||||||
Db        840 CACCCCCCCCCCACCCCCCCCCGCCCCCCCCCCCCCCCCCACCCCCCCACACCCACCCCCC 899

Qy        246 CACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCCC 305
              | || ||  || || | || ||  |  || |  |   | | ||| | ||  || ||||||
Db        900 CCCCCCCCCCACCCCCCCCCCACCCCCCCCCCGCCCCCACCCCCCCCCCCCCCCCACCCC 959

Qy        306 CACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAACT 365
              | |||| | |  |      || || | || |  ||  |  | ||| || || || ||  |
Db        960 CCCCCACCCCCCCGCACACCACCCCCCCCACACCACCCCCCCCCCCCCCCCCCCCCCCCC 1019

Qy        366 CACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCGG 425
              | |||| | |||||||  | || ||  |  || |  | || ||||| || || ||  |
Db       1020 CCCCCACCCCACCCCCACCCCCCCCCCCCCCCCCCCCCCCACCCCCGCCCCCCCCCCCCCC 1079

Qy        426 CGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACCA 485
              | | |||  ||   ||  | || |   | | ||  ||     | ||| | ||||||
Db       1080 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCCCACCCCCCCCCCCCCCC 1139

Qy        486 CCGCCACACCCACCACCTCCCCGCCACCAGCATCGCC 522
              || || | ||| || || |||| || ||  |  | ||
Db       1140 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 1176
```

RESULT 14

BM908878
LOCUS BM908878 1303 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6611187 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5431479 5', mRNA sequence.
ACCESSION BM908878
VERSION BM908878.1 GI:19359257
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1303)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1903 row: m column: 16
High quality sequence stop: 135.
FEATURES Location/Qualifiers
source 1. .1303
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5431479"
/clone_lib="NIH_MGC_98"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 154 a 913 c 55 g 105 t 76 others
ORIGIN

Query Match 29.1%; Score 155.4; DB 14; Length 1303;
Best Local Similarity 53.4%; Pred. No. 9.7e-17;
Matches 282; Conservative 0; Mismatches 246; Indels 0; Gaps 0;

```
Qy          5 CCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCA 64
              || ||| |  | ||  |    ||  | ||| | ||  || || | ||| || || | ||
Db        297 CCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCNNCCNCCCCCCCCCCCCCCCCCCCC 356

Qy         65 CCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACG 124
              ||| ||  ||| |||  | || || || | ||  |    ||  |||||  | |||||  |
Db        357 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCNCCC 416

Qy        125 CACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCG 184
              | |  | ||   || |||  | ||  |          |  || || | ||  | || ||  |
Db        417 CCCCCCCCCCNNCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCNCCCCCCCCC 476

Qy        185 ACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCC 244
              | || |  |||   ||  | ||  |  | || |  | ||  ||||||  |  | |||||
Db        477 CCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCC 536

Qy        245 TCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCC 304
              |  | ||  |  || | ||  |   | || |  | ||  ||| | ||  || | |||
Db        537 CCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 596

Qy        305 CCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAAC 364
              | |||  | |  | ||| || | ||     ||| |  | ||| || || |  | |  |
Db        597 NCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCNCNNCCCCC 656
```

```
Db        141 CCCCCCCCCCCCCCCCCCCCCCCGCCCMCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCMCC 200

Qy        183 CGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCC 242

Db        201 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCMCCCCCC 260

Qy        243 CCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCAC 302

Db        261 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 320

Qy        303 CCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCA 362

Db        321 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACMCCCCCCCCCCCCCCCCCCCCCCC 380

Qy        363 ACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTT 422

Db        381 CCCCMMCCCCCCCCCCCCCCCCCCCACCCCCMCCCCCACCCACCCCCCCCCCCACCCCCCC 440

Qy        423 CGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCA 482

Db        441 CCMCCCACCCACCCCCCMCCCCCCCCCCCCCCCCCCACMCCCCCCCCCCCCCCCCCCCCCCM 500

Qy        483 CCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCC 522

Db        501 CCMCCCCCCCCCCCCCCACCCCCCCCCCCMCCCCCCCCCC 540


RESULT 12
BQ939440/c
LOCUS       BQ939440               1089 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_8929085 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6484345
            5', mRNA sequence.
ACCESSION   BQ939440
VERSION     BQ939440.1  GI:22354918
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1089)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2670  row: g  column: 02
            High quality sequence stop: 327.
FEATURES             Location/Qualifiers
     source          1. .1089
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6484345"
                     /clone_lib="NIH_MGC_40"
                     /tissue_type="carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed  by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California,  Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life  Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT       73 a     59 c    712 g     81 t    164 others
ORIGIN
```

```
  Query Match           29.1%;  Score 155.6;  DB 14;  Length 1089;
  Best Local Similarity   52.5%;  Pred. No. 9.1e-17;
  Matches  272;  Conservative    0;  Mismatches  246;  Indels    0;  Gaps    0;

Qy          5 CCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCA 64

Db       1034 CCCCCCCCNCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCNCCCCNCCCCCCNCCCCCCCC 975

Qy         65 CCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACG 124

Db        974 CCCCCCCCCCNCCCCCCNCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 915

Qy        125 CACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCG 184

Db        914 CCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCNCCCCNCCCCCCCCCCNNCNNCCCCNNCC 855

Qy        185 ACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCC 244

Db        854 CCCCCCCCCCCNCCCCCCCNCCCCNNCCCCCCCCCCCCCCCCCNNCNCNNNCCNCCCNCCN 795

Qy        245 TCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCC 304

Db        794 NCNNNNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCC 735

Qy        305 CCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAAC 364

Db        734 CCCCCCCCNNNCNNNNCCCCCCCCCCCCCCCNNCNCNNNNCCCCCCCCCCCCCCCCCCCC 675

Qy        365 TCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCG 424

Db        674 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 615

Qy        425 GCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACC 484

Db        614 CCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCC 555

Qy        485 ACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCC 522

Db        554 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 517


RESULT 13
BG067997
LOCUS       BG067997               1228 bp    mRNA    linear   EST 26-JAN-2001
DEFINITION  H3060F02-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
            H3060F02 3', mRNA sequence.
ACCESSION   BG067997
VERSION     BG067997.1  GI:12550566
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
            ,T.S., Carter,M.G. and Ko,M.S.H.
  TITLE     Verification and initial annotation of NIA mouse 15K cDNA clone set
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: H3060F02-5
            Contact: George J. Kargul
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            This clone set has been freely distributed to the community. Please
            visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
            Plate: H3060  row: F  column: 02
            Seq primer: -21M13 Forward
            High quality sequence stop: 1228
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="Mus musculus"
```

```
RESULT 10
BG786340/c
LOCUS       BG786340                688 bp    mRNA    linear   EST 20-MAY-2001
DEFINITION  SEAUMC006297 Sea urchin primary mesenchyme cell cDNA library
            Strongylocentrotus purpuratus cDNA clone PC_0028_A2_H09_MR 5', mRNA
            sequence.
ACCESSION   BG786340
VERSION     BG786340.1  GI:14157353
KEYWORDS    EST.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 688)
  AUTHORS   Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
            Ettensohn,C.A.
  TITLE     A large scale analysis of mRNAs expressed by primary mesenchyme
            cells of the sea urchin embryo
  JOURNAL   Development 128 (13), 2615-2627 (2001)
  MEDLINE   21384984
COMMENT     Contact: Ettensohn CA
            Dept. Biol. Sci.
            Carnegie Mellon University
            4400 Fifth Avenue, Pittsburgh, PA 15213, USA
            Tel: +1 412 268 5849
            Email: ettensohn@andrew.cmu.edu.
FEATURES             Location/Qualifiers
     source          1. .688
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="PC_0028_A2_H09_MR"
                     /clone_lib="Sea urchin primary mesenchyme cell cDNA
                     library"
                     /tissue_type="embryo"
                     /cell_type="primary mesenchyme cells"
                     /lab_host="E.coli"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; oligo
                     dt priming from poly A+ RNA, directionally cloned"
BASE COUNT        9 a      0 c    596 g     53 t     30 others
ORIGIN

  Query Match             29.2%;  Score 155.8;  DB 12;  Length 688;
  Best Local Similarity   55.7%;  Pred. No. 8.6e-17;
  Matches  291;  Conservative    0;  Mismatches  227;  Indels    4;  Gaps    1;

Qy          5 CCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCA 64
              || ||| |  ||||  |    ||  | ||| | || |||||| | ||| || || | |||
Db        566 CCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCA 507

Qy         65 CCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACG 124
              ||| ||  ||| |||  | || || || | ||  ||   ||  | |||   |||||| |
Db        506 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCACCCCCCCCC 447

Qy        125 CACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTG-- 182
              | |  | || | || |||  | ||| || |    || || ||   ||  |  | ||
Db        446 CCCCCCCCCCCCCCCCCCCCCCCCACCACANCCCCAACCNCCNNCCCNCCCNCCCCNNAC 387

Qy        183 --CGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
                | | ||  | |||| ||  |   |  |  | || | || ||  |||||| |  | |
Db        386 CCCCCCCCCCACCCCCCACACACANCNCCCCCCCCCCCCCACCCCCNCCCCCCCCCCCCC 327

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              |||  ||  ||   |  || |||  |   |  |  | || ||| ||   ||  || | |
Db        326 CCCACCAAAACACACCCCCCCACNCCCACCCCCACCCCCCACNCCANCCCNCCCCCCCCC 267

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              ||||  |||  | | ||  |  || || | ||  | |||||  | ||| || || ||||
Db        266 CCCCCNCCCCCCCCCCACCCCCCCCCCCCCCCCCNCACCCCACCCCCCCCCCCCCCCCAC 207

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              || |   |||| ||  |||||| | ||  |  |  || |  | || || || || || ||
```

```
Db        206 CACCCACCCCACCANCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCC 147

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
                |  | | |||  |||    |  |  | || |  ||  | ||  || |  | |||   |
Db        146 CNCCCCCCCCCCCCCCACCNCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCA 87

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCC 522
              |   ||  || | ||| || || |||| || ||  |  | ||
Db         86 CNAAACNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 45


RESULT 11
CNS046XY
LOCUS       CNS046XY                638 bp    DNA     linear   GSS 18-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            086D07 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL277279
VERSION     AL277279.1  GI:8011482
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 638)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 638)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 638)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .638
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="086D07"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG086CB04SP1~end :
                     PUC-Ori"
BASE COUNT       59 a    513 c     16 g     21 t     29 others
ORIGIN

  Query Match             29.1%;  Score 155.6;  DB 17;  Length 638;
  Best Local Similarity   52.9%;  Pred. No. 9.3e-17;
  Matches  275;  Conservative   23;  Mismatches  222;  Indels    0;  Gaps    0;

Qy          3 AACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCAC 62
              | |  | | || ||||        ||| | ||| | || || |||   :::  || | | |
Db         21 ATCAGCACCCACCACCCCACCCCCCGCCTCCCACCCCCCCAACCCTGSMMCACAACMCCC 80

Qy         63 CACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCA 122
              | ||| ||  ||| ||| : :|| :| |: ::||   |  |||  |:||| | : || |:
Db         81 CCCCCCCCCCCCCCCCCCMAMCCCMCCCMCMMCCCCACMCACCCCCMCCCCCCMACCACM 140

Qy        123 CGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTG 182
              | | |  | || | || |||    ||: |       |  || || ||||  | || |:
```

```
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM1986  row: k  column: 12
          High quality sequence stop: 218.
FEATURES             Location/Qualifiers
     source          1. .1161
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5474051"
                     /clone_lib="NIH_MGC_41"
                     /tissue_type="amelanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT       53 a    844 c    107 g     88 t     69 others
ORIGIN

  Query Match             29.3%;  Score 156.6;  DB 14;  Length 1161;
  Best Local Similarity   54.4%;  Pred. No. 6.2e-17;
  Matches  282;  Conservative    0;  Mismatches  236;  Indels    0;  Gaps    0;

Qy          5 CCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCA 64
              || ||| |  | ||        ||  | ||| | || ||| || | ||| || || | ||
Db        312 CCCCCCCCCCCCCCCNNNCCCCCCCCCCCCCCCNCCCCCCNCCCCCCCCCCCCCCCCCCC 371

Qy         65 CCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACG 124
              ||| ||  ||| |||  | |  || || | ||  ||    ||  ||||| | |||||| |
Db        372 CCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCNNCCCCCCCCCCCCCCCCCCCCC 431

Qy        125 CACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCG 184
              | |  | || |  | |||  | ||  |       |  || || | ||  | || ||  |
Db        432 CCCCNCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCC 491

Qy        185 ACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCC 244
              | ||  |  |||   ||  | ||  |  | || |  | ||  |||||| |  | |||||
Db        492 CCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 551

Qy        245 TCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCC 304
              | || ||  |   | | || |   | || |  | | || ||| | ||  || | |||
Db        552 CCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 611

Qy        305 CCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAAC 364
              ||'|||  | |  |  |  || ||    |   ||| |  | ||| || || || ||
Db        612 CCCCCCCCCCCCCCCCCCCCCCCCCCCNNNCNNCCCCCCCCCCCCCNCCCCCCCCCCCCN 671

Qy        365 TCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCG 424
              |||  | |||||| | || ||  |  || |  | || || || || || ||  |
Db        672 NNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCC 731

Qy        425 GCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACC 484
              | | |||  ||   ||  |  | || |  ||  | ||  || |  | ||| | ||| ||
Db        732 NCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 791

Qy        485 ACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCC 522
              || || | ||  || || |||| || ||  |  | ||
Db        792 CCCCCCCCCCCNNCCCCCCCCCCCCCCCCCCCCCCNCC 829


RESULT 9
BI416700/c
LOCUS       BI416700                842 bp    mRNA    linear   EST 15-AUG-2001
DEFINITION  hasp001xp03f Heterobasidion annosum - Scots pine infection stage
            subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
            annosum cDNA clone hasp001xp03f, mRNA sequence.
ACCESSION   BI416700
VERSION     BI416700.1  GI:15187723
KEYWORDS    EST.
SOURCE      Pinus sylvestris/Heterobasidion annosum.
  ORGANISM  Pinus sylvestris/Heterobasidion annosum
            Eukaryota; mixed EST libraries.
REFERENCE   1  (bases 1 to 842)
  AUTHORS   Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
  TITLE     Expressed sequence tags of randomly selected cDNA clones from the
            interaction of the root rot fungus (Heterobasidion annosum) with
            seedling roots of Scots pine (Pinus sylvestris)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Fred O. Asiegbu
            Dept. of Forest Mycology & Pathology
            Swedish University of Agriculture, Box 7026,S-750 07, Uppsala,
            Sweden
            Tel: +46 18 67 15 98
            Fax: +46 18 30 92 45
            Email: Fred.Asiegbu@mykopat.slu.se
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .842
                     /organism="Pinus sylvestris/Heterobasidion annosum"
                     /db_xref="taxon:169015"
                     /clone="hasp001xp03f"
                     /clone_lib="Heterobasidion annosum - Scots pine infection
                     stage subtraction cDNA library (hasp)"
                     /dev_stage="Seedling roots of scots pine were infected for
                     6 days with H. annosum"
                     /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive
                     hybridization cDNA library was constructed from  scots
                     pine roots infected for 6-days with mycelia of
                     Heterobasidon  annosum (FP5)."
BASE COUNT       17 a     14 c    654 g     50 t    107 others
ORIGIN

  Query Match             29.3%;  Score 156.4;  DB 13;  Length 842;
  Best Local Similarity   54.1%;  Pred. No. 6.8e-17;
  Matches  280;  Conservative    0;  Mismatches  238;  Indels    0;  Gaps    0;

Qy          5 CCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCA 64
              || | | |  |||    |    ||  |  || | | |||| || | |||  || || | ||
Db        730 CCCCACCCCCCACACCCCCCCCCCCCCCACCCCCCNACCCCCCCCCCCCCCCCCCCCCCC 671

Qy         65 CCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACG 124
              || | | ||| |||  | |  || || | ||  ||    ||  ||||| |||||||| |
Db        670 ACCCCAACCCCCCCCCCCCCCNCCCCCCCCCNCCCACCCCCCCCCCCCCCCCACCCCCCCC 611

Qy        125 CACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCG 184
              | |  | || | || |||| | || |||       |  || || | ||  |||| |    |
Db        610 CCCCCCCCCCCCCCCCCCCANCCCCCACACCCCCCCCCCCCCCCCCCCCCCACCCCNCCCC 551

Qy        185 ACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCC 244
                 ||  |  ||    ||  | |    |  | || || | ||  | |||| |  | ||| |
Db        550 CACCCCACCCCCACCCCCCCCCCACCCCCCCCCACACCCCCNACNCCCCCCCCCCCCCAC 491

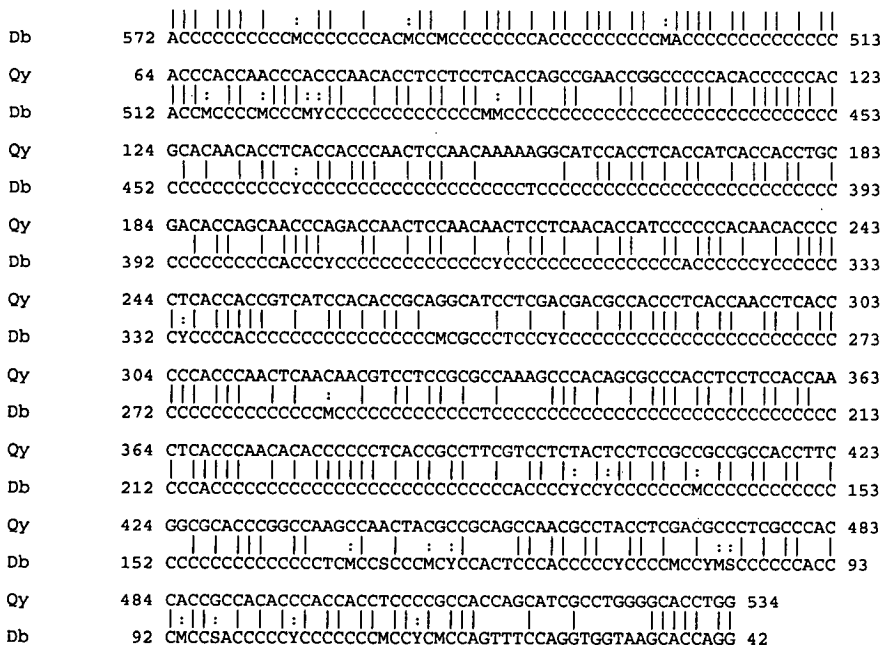
Qy        245 TCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCC 304
              | || ||  |  || | ||  |   | || |  | || || ||| | ||  || | |||
Db        490 CCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCC 431

Qy        305 CCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAAC 364
              || |||  | | ||   |  || || | || || ||| |   | ||| || ||  | ||  |
Db        430 CCNCCCCCCCCCACCCCCCCCCCCCCCNCCCAACCCCCCCCCCCCCCCCCCCCNCCCCCCC 371

Qy        365 TCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCG 424
              | |||  | | |||||| | || ||  |  || |  | || ||    || || ||  |
Db        370 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCNNNCCCCCCCCCCCC 311

Qy        425 GCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACC 484
              | |||||  |   ||  | | |  | || | ||   | |   ||  | ||| ||
Db        310 CCCCACCCCNCNNNNCCCNCCCCCCNNCCCCCCCCCNCCCCNCCCCCNCCCNNCCCCCCCC 251

Qy        485 ACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCC 522
              ||  |   ||  |  || |||| || ||  |  | ||
Db        250 CCCCNCNNCCCNCCNCCCCCCCCCCCCCCCCCCCCCCC 213
```



```
Db        572 ACCCCCCCCCCCMCCCCCCCACMCCMCCCCCCCCCACCCCCCCCCCCMACCCCCCCCCCCCCCC 513

Qy         64 ACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCAC 123

Db        512 ACCMCCCCMCCCMYCCCCCCCCCCCCCCCMMCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 453

Qy        124 GCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGC 183

Db        452 CCCCCCCCCCCYCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCC 393

Qy        184 GACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCC 243

Db        392 CCCCCCCCCCACCCYCCCCCCCCCCCCCCCYCCCCCCCCCCCCCCCACCCCCCYCCCCCC 333

Qy        244 CTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACC 303

Db        332 CYCCCCACCCCCCCCCCCCCCCCCMCGCCCTCCCYCCCCCCCCCCCCCCCCCCCCCCCCC 273

Qy        304 CCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAA 363

Db        272 CCCCCCCCCCCCCCMCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 213

Qy        364 CTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTC 423

Db        212 CCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCYCCYCCCCCCCMCCCCCCCCCCC 153

Qy        424 GGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCAC 483

Db        152 CCCCCCCCCCCCCCTCMCCSCCCMCYCCACTCCCACCCCCYCCCCMCCYMSCCCCCCACC 93

Qy        484 CACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534

Db         92 CMCCSACCCCCYCCCCCCCMCCYCMCCAGTTTCCAGGTGGTAAGCACCAGG 42
```

RESULT 7
AZ186654
LOCUS AZ186654 897 bp DNA linear GSS 30-AUG-2000
DEFINITION SP_1007_B1_D06_SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=1007 Col=11 Row=H, DNA sequence.
ACCESSION AZ186654
VERSION AZ186654.1 GI:8369832
KEYWORDS GSS.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.
REFERENCE 1 (bases 1 to 897)
AUTHORS Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and Hood,L.
TITLE A sea urchin genome project: Sequence scan, virtual map, and additional resources
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
MEDLINE 20402566
COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 1007 row: H column: 11
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 897.
FEATURES Location/Qualifiers
source 1. .897
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="Plate=1007 Col=11 Row=H"
/clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli DH10B"
BASE COUNT 31 a 646 c 34 g 54 t 132 others
ORIGIN

Query Match 29.3%; Score 156.6; DB 17; Length 897;
Best Local Similarity 53.4%; Pred. No. 6.3e-17;
Matches 276; Conservative 0; Mismatches 241; Indels 0; Gaps 0;


```
Qy          5 CCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCA 64

Db        381 CCCCCCCCCCCNCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 440

Qy         65 CCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACG 124

Db        441 CCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCNNCCCCCNNCCCCCCCCCCCCCCC 500

Qy        125 CACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCG 184

Db        501 CCCCCCCCCCNCNNCNCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 560

Qy        185 ACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCC 244

Db        561 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNNC 620

Qy        245 TCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCC 304

Db        621 CCCCNCCNCCCCCCCCCCCCCCCCNCCCCCCCCCNCCCCCCCCNCCCNCCCCCCCCCCCC 680

Qy        305 CCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAAC 364

Db        681 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCNNCCCNNCCCNCCCCCCTCCNCCCNCNNN 740

Qy        365 TCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCG 424

Db        741 NCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCC 800

Qy        425 GCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACC 484

Db        801 CCCCCCCNCCCCCGCCCNCCCCCCCCCCCCCCCCCCNCCCCCCNCCNNCNCCCCCCCCCCC 860

Qy        485 ACCGCCACACCCACCACCTCCCCGCCACCAGCATCGC 521

Db        861 CCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCNCCC 897
```

RESULT 8
BM912505
LOCUS BM912505 1161 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6612272 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5474051 5', mRNA sequence.
ACCESSION BM912505
VERSION BM912505.1 GI:19362884
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1161)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be

```
Qy        305 CCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAAC 364
              || |||  | |  |  |  || || | ||  | ||| |  | ||| ||  | || ||  |
Db        467 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCTCCCCCCCCC 408

Qy        365 TCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCG 424
              | |||  | | ||||||| | || ||  |  || |  | || || || || || ||  |  |
Db        407 CCCCCCCCCTCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCTCCCCNCCCC 348

Qy        425 GCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACC 484
              | | ||| ||  ||  | | || |  ||  | ||  || |  | ||| | ||| ||
Db        347 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 288

Qy        485 ACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCC 522
              |  || | |||  | || |||| || ||  |  | ||
Db        287 CNCCCCCCTCCCCTCCCCCCCCCCCCCCCCCCCCCCC 250


RESULT 5
BI416659/c
LOCUS       BI416659                806 bp    mRNA    linear   EST 15-AUG-2001
DEFINITION  hasp001xm18f Heterobasidion annosum - Scots pine infection stage
            subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
            annosum cDNA clone hasp001xm18f, mRNA sequence.
ACCESSION   BI416659
VERSION     BI416659.1  GI:15187682
KEYWORDS    EST.
SOURCE      Pinus sylvestris/Heterobasidion annosum.
  ORGANISM  Pinus sylvestris/Heterobasidion annosum
            Eukaryota; mixed EST libraries.
REFERENCE   1  (bases 1 to 806)
  AUTHORS   Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
  TITLE     Expressed sequence tags of randomly selected cDNA clones from the
            interaction of the root rot fungus (Heterobasidion annosum) with
            seedling roots of Scots pine (Pinus sylvestris)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Fred O. Asiegbu
            Dept. of Forest Mycology & Pathology
            Swedish University of Agriculture, Box 7026,S-750 07, Uppsala,
            Sweden
            Tel: +46 18 67 15 98
            Fax: +46 18 30 92 45
            Email: Fred.Asiegbu@mykopat.slu.se
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .806
                     /organism="Pinus sylvestris/Heterobasidion annosum"
                     /db_xref="taxon:169015"
                     /clone="hasp001xm18f"
                     /clone_lib="Heterobasidion annosum - Scots pine infection
                     stage subtraction cDNA library (hasp)"
                     /dev_stage="Seedling roots of scots pine were infected for
                     6 days with H. annosum"
                     /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive
                     hybridization cDNA library was constructed from  scots
                     pine roots infected for 6-days with mycelia of
                     Heterobasidon  annosum (FP5)."
BASE COUNT       13 a     15 c    676 g     35 t     67 others
ORIGIN

  Query Match           29.6%;  Score 158;  DB 13;  Length 806;
  Best Local Similarity   56.0%;  Pred. No. 3.7e-17;
  Matches  290;  Conservative    0;  Mismatches  228;  Indels    0;  Gaps    0;

Qy          5 CCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCA 64
              || ||| |  | ||  |     ||  | ||| | || |||  | | ||| || || | ||
Db        707 CCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCC 648

Qy         65 CCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACG 124
              ||| ||  ||| |||  | || || || | ||  ||   ||  ||||| | |||||  |
Db        647 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCACCC 588

Qy        125 CACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCG 184
```

```
              | | . | || | || |||  | |||||        |  || || | ||  | |  ||  |
Db        587 CCCCNCCCCCCCCCCCCCCCCCCCCCAACCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCC 528

Qy        185 ACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCC 244
              | || |   ||   ||  | ||  |  | || |  | ||  |||||| | |   ||||
Db        527 CCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCAACACCCC 468

Qy        245 TCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCC 304
              | || ||  |  || | || |   ||  | | ||  | ||| || | ||  || | |||
Db        467 CCCCCCCCCCCCCCCCCCCCCCCCCCACACCCCACACCCCCAACCCCCCCCCCCCCCCCC 408

Qy        305 CCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAAC 364
              || |||| | |  |  | || || | ||  | ||| |  | ||| || || || ||  |
Db        407 CCCCCCACCCCCCCNCCCCCCCCCCCCCCCCCAACCCCCNACCCCCCCCCCCCCCCCCCC 348

Qy        365 TCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCG 424
              | |||  | | |||||| | || ||  |  || |  | || || || || || ||  |  |
Db        347 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCC 288

Qy        425 GCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACC 484
              | |  ||  ||   ||  |  | || |  ||  | ||  || |  | ||| | ||| ||
Db        287 CCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCC 228

Qy        485 ACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCC 522
              || || | ||| || || |||| || ||  |  | ||
Db        227 CCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCC 190


RESULT 6
AL570062/c
LOCUS       AL570062                715 bp    mRNA    linear   EST 16-FEB-2001
DEFINITION  AL570062 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI006YC11 3
            prime, mRNA sequence.
ACCESSION   AL570062
VERSION     AL570062.1  GI:12926016
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 715)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .715
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DI006YC11"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT       35 a     31 c    574 g     37 t     38 others
ORIGIN

  Query Match           29.4%;  Score 157;  DB 9;  Length 715;
  Best Local Similarity   52.7%;  Pred. No. 5.4e-17;
  Matches  280;  Conservative   23;  Mismatches  228;  Indels    0;  Gaps    0;

Qy          4 ACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACC 63
```

```
LOCUS       BI416575                931 bp    mRNA    linear   EST 15-AUG-2001
DEFINITION  hasp001xg15f Heterobasidion annosum - Scots pine infection stage
            subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
            annosum cDNA clone hasp001xg15f, mRNA sequence.
ACCESSION   BI416575
VERSION     BI416575.1  GI:15187598
KEYWORDS    EST.
SOURCE      Pinus sylvestris/Heterobasidion annosum.
  ORGANISM  Pinus sylvestris/Heterobasidion annosum
            Eukaryota; mixed EST libraries.
REFERENCE   1  (bases 1 to 931)
  AUTHORS   Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
  TITLE     Expressed sequence tags of randomly selected cDNA clones from the
            interaction of the root rot fungus (Heterobasidion annosum) with
            seedling roots of Scots pine (Pinus sylvestris)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Fred O. Asiegbu
            Dept. of Forest Mycology & Pathology
            Swedish University of Agriculture, Box 7026,S-750 07, Uppsala,
            Sweden
            Tel: +46 18 67 15 98
            Fax: +46 18 30 92 45
            Email: Fred.Asiegbu@mykopat.slu.se
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .931
                     /organism="Pinus sylvestris/Heterobasidion annosum"
                     /db_xref="taxon:169015"
                     /clone="hasp001xg15f"
                     /clone_lib="Heterobasidion annosum - Scots pine infection
                     stage subtraction cDNA library (hasp)"
                     /dev_stage="Seedling roots of scots pine were infected for
                     6 days with H. annosum"
                     /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive
                     hybridization cDNA library was constructed from  scots
                     pine roots infected for 6-days with mycelia of
                     Heterobasidon  annosum (FP5)."
BASE COUNT        9 a      7 c    884 g     15 t     16 others
ORIGIN

  Query Match             29.9%;  Score 159.6;  DB 13;  Length 931;
  Best Local Similarity   56.2%;  Pred. No. 2e-17;
  Matches  291;  Conservative    0;  Mismatches  227;  Indels    0;  Gaps    0;

Qy          5 CCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCA 64
              || ||| |  | ||  |  | ||  | ||| | || | | || | ||| || || ||||
Db        680 CCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCACCC 621

Qy         65 CCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACG 124
              ||| ||  ||| |||  | || || || | ||  ||   ||  ||| | | |||||| |
Db        620 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCACCCCCCCCCCCC 561

Qy        125 CACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCG 184
              | |  | || | || |||| | ||  |       |  || || | ||  | || ||  |
Db        560 CCCCCCCCCCCCCCCCCACCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 501

Qy        185 ACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCC 244
              | ||  |  |||   ||  | ||  |  | || |  | ||  |||||| |  |  ||||
Db        500 CCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNNCCCC 441

Qy        245 TCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCC 304
              | || ||  |  || | ||  |   | || |  |  | || ||| | ||  || | |||
Db        440 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 381

Qy        305 CCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAAC 364
              || |||  | |  |  | || || | ||     ||| |  | |||  || || || ||  |
Db        380 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCC 321

Qy        365 TCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCG 424
              | |||  | | ||||||  | |  ||  |  || |  | || || || || || ||  |
Db        320 CCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 261

Qy        425 GCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACC 484
              | | |||  ||   ||  |  | || |  ||  | ||  || |  |||| || ||| ||
Db        260 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCTTCCCCCCCC 201

Qy        485 ACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCC 522
              || || ||||| || || |||| |  ||  |  | ||
Db        200 CCCCCCCCACCCCCCCCCNCCCCCCNCCCCCCCCCCC 163


RESULT 4
BI416650/c
LOCUS       BI416650                1015 bp    mRNA    linear   EST 15-AUG-2001
DEFINITION  hasp001xm01f Heterobasidion annosum - Scots pine infection stage
            subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
            annosum cDNA clone hasp001xm01f, mRNA sequence.
ACCESSION   BI416650
VERSION     BI416650.1  GI:15187673
KEYWORDS    EST.
SOURCE      Pinus sylvestris/Heterobasidion annosum.
  ORGANISM  Pinus sylvestris/Heterobasidion annosum
            Eukaryota; mixed EST libraries.
REFERENCE   1  (bases 1 to 1015)
  AUTHORS   Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
  TITLE     Expressed sequence tags of randomly selected cDNA clones from the
            interaction of the root rot fungus (Heterobasidion annosum) with
            seedling roots of Scots pine (Pinus sylvestris)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Fred O. Asiegbu
            Dept. of Forest Mycology & Pathology
            Swedish University of Agriculture, Box 7026,S-750 07, Uppsala,
            Sweden
            Tel: +46 18 67 15 98
            Fax: +46 18 30 92 45
            Email: Fred.Asiegbu@mykopat.slu.se
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .1015
                     /organism="Pinus sylvestris/Heterobasidion annosum"
                     /db_xref="taxon:169015"
                     /clone="hasp001xm01f"
                     /clone_lib="Heterobasidion annosum - Scots pine infection
                     stage subtraction cDNA library (hasp)"
                     /dev_stage="Seedling roots of scots pine were infected for
                     6 days with H. annosum"
                     /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive
                     hybridization cDNA library was constructed from  scots
                     pine roots infected for 6-days with mycelia of
                     Heterobasidon  annosum (FP5)."
BASE COUNT       21 a      6 c    961 g     16 t     11 others
ORIGIN

  Query Match             29.8%;  Score 159.2;  DB 13;  Length 1015;
  Best Local Similarity   56.6%;  Pred. No. 2.3e-17;
  Matches  293;  Conservative    0;  Mismatches  225;  Indels    0;  Gaps    0;

Qy          5 CCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCA 64
              || ||| |  | ||  |     ||  | ||| | || ||| || | |||| || || ||||
Db        767 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCC 708

Qy         65 CCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACG 124
              ||| ||  ||| |||  | || || || | ||  ||   ||  ||||| | |||||| |
Db        707 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 648

Qy        125 CACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCG 184
              | |  | || | || |||  | ||  |       |  || || | ||  | || ||  |
Db        647 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 588

Qy        185 ACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCC 244
              | ||  |  |||   ||  |||| ||  | || |  | ||  |||||| |     |||||
Db        587 TCCCCCCCCCCCCCCCCCCCCCTCCCACCCCCCCCCCCCCCCCCCCCCCCCCCTACCCCCC 528

Qy        245 TCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCC 304
              | |||||  |  || | || |   |  || |  |  | |||||| |||| || | |||
Db        527 CCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCACCCCCCCCCC 468
```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5480305"
/clone_lib="NIH_MGC_113"
/lab_host="DH10B (phage-resistant)"
/note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 73 a 824 c 62 g 234 t 43 others
ORIGIN

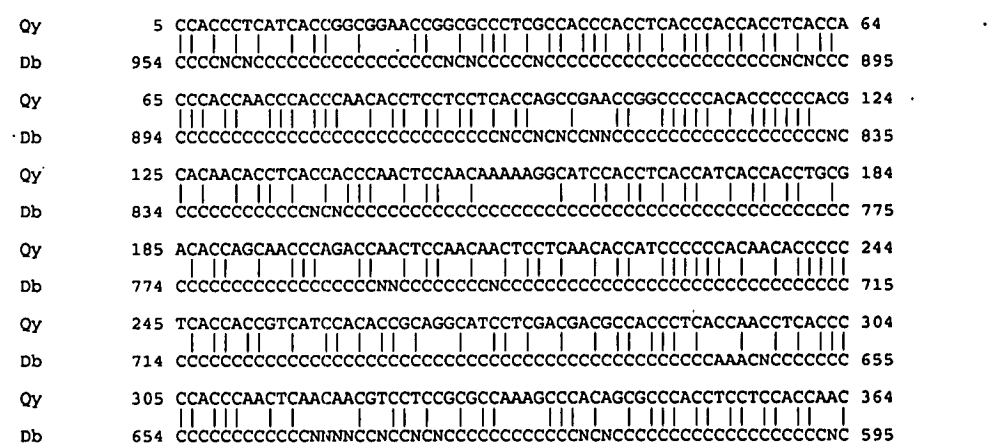
Query Match 30.0%; Score 160; DB 14; Length 1236;
Best Local Similarity 54.8%; Pred. No. 1.7e-17;
Matches 289; Conservative 0; Mismatches 238; Indels 0; Gaps 0;



RESULT 2
BM913931/c
LOCUS BM913931 1243 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6608063 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5478793 5', mRNA sequence.
ACCESSION BM913931
VERSION BM913931.1 GI:19364310
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1243)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1999 row: a column: 02
High quality sequence stop: 369.
FEATURES Location/Qualifiers
source 1. .1243
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5478793"
/clone_lib="NIH_MGC_98"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 47 a 193 c 776 g 116 t 111 others
ORIGIN

Query Match 30.0%; Score 160; DB 14; Length 1243;
Best Local Similarity 53.5%; Pred. No. 1.7e-17;
Matches 277; Conservative 0; Mismatches 241; Indels 0; Gaps 0;



RESULT 3
BI416575/c

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 684.535 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_5143_5676
Perfect score: 534
Sequence: 1 ggaaccaccctcatcaccgg.....gcatcgcctggggcacctgg 534

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	160	30.0	1236	14	BM914619	BM914619 AGENCOURT
c	2	160	30.0	1243	14	BM913931	BM913931 AGENCOURT
c	3	159.6	29.9	931	13	BI416575	BI416575 hasp001xg
c	4	159.2	29.8	1015	13	BI416650	BI416650 hasp001xm
c	5	158	29.6	806	13	BI416659	BI416659 hasp001xm
c	6	157	29.4	715	9	AL570062	AL570062 AL570062
	7	156.6	29.3	897	17	AZ186654	AZ186654 SP_1007_B
	8	156.6	29.3	1161	14	BM912505	BM912505 AGENCOURT
c	9	156.4	29.3	842	13	BI416700	BI416700 hasp001xp
c	10	155.8	29.2	688	12	BG786340	BG786340 SEAUMC006
	11	155.6	29.1	638	17	CNS046XY	AL277279 Tetraodon
c	12	155.6	29.1	1089	14	BQ939440	BQ939440 AGENCOURT
	13	155.4	29.1	1228	12	BG067997	BG067997 H3060F02-
	14	155.4	29.1	1303	14	BM908878	BM908878 AGENCOURT
c	15	155	29.0	1197	13	BI416470	BI416470 hasp001xa
c	16	154.8	29.0	660	17	AG089410	AG089410 Pan trogl
c	17	154.8	29.0	1005	17	BH148610	BH148610 ENTPZ67TR
	18	154.6	29.0	703	17	AZ185816	AZ185816 SP_1006_A
c	19	154.6	29.0	1011	17	AG062256	AG062256 Pan trogl
c	20	154.6	29.0	1089	14	BQ930965	BQ930965 AGENCOURT
c	21	154.6	29.0	1138	17	AG152606	AG152606 Pan trogl
c	22	154.4	28.9	703	12	BG786294	BG786294 SEAUMC006
c	23	154.4	28.9	863	13	BI416699	BI416699 hasp001xp
c	24	154	28.8	748	17	AZ187748	AZ187748 SP_1009_B
c	25	154	28.8	765	17	AG043737	AG043737 Pan trogl
c	26	154	28.8	843	17	AZ200896	AZ200896 SP_1011_A
c	27	153.8	28.8	1047	14	BQ894173	BQ894173 AGENCOURT
c	28	153.6	28.8	1088	12	BF256580	BF256580 HVSMEf001
	29	153.2	28.7	877	17	AZ185000	AZ185000 SP_1004_A
c	30	153.2	28.7	993	17	AG081720	AG081720 Pan trogl
c	31	153	28.7	1184	14	BM911719	BM911719 AGENCOURT
c	32	152.8	28.6	903	17	BH136778	BH136778 ENTPA49TF
	33	152.8	28.6	1089	13	BM468944	BM468944 AGENCOURT
c	34	152.6	28.6	817	17	AZ196702	AZ196702 SP_1032_B
c	35	152.6	28.6	897	13	BI416517	BI416517 hasp001xd
	36	152.4	28.5	707	17	AZ193616	AZ193616 SP_1023_B
c	37	152.4	28.5	780	10	BE641449	BE641449 Cri2_3_E2
c	38	152.4	28.5	1222	17	BH146899	BH146899 ENTQM47TF
c	39	152.2	28.5	803	13	BI416474	BI416474 hasp001xa
c	40	152.2	28.5	993	17	AG033032	AG033032 Pan trogl
	41	152	28.5	672	17	AG133158	AG133158 Pan trogl
c	42	152	28.5	888	17	AZ189327	AZ189327 SP_1014_A
c	43	152	28.5	1052	17	AQ740804	AQ740804 HS_5508_A
c	44	152	28.5	1478	14	BM911600	BM911600 AGENCOURT
c	45	151.8	28.4	540	17	CNS0103K	AL098474 Drosophil

ALIGNMENTS

RESULT 1
BM914619
LOCUS BM914619 1236 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6615286 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:5480305 5', mRNA sequence.
ACCESSION BM914619
VERSION BM914619.1 GI:19364998
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1236)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Mark Watson
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2002 row: p column: 02
High quality sequence stop: 147.
FEATURES Location/Qualifiers
source 1. .1236

```
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-1

  Query Match             36.8%;  Score 196.4;  DB 9;  Length 15872;
  Best Local Similarity   60.5%;  Pred. No. 6.6e-35;
  Matches  323;  Conservative    0;  Mismatches  211;  Indels    0;  Gaps    0;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||| ||||||| || ||||| | |||||   | | |||| ||| |||||||
Db       7337 GGCACCGTCCTCGTCACCGGGGGCACCGGAGGCCTCGGAGCGCTCCTCGCCCGCCACCTC 7396

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
                |  ||       |  |  || || ||    |||||   |||  || || |||||
Db       7397 GTGGAGCGTCACGGGGTGCGCCGGCTGCTGCTGGTGAGCCGCCGCGGGGCCGACGCCCCC 7456

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
               |||     ||||   |    | |||         |||   .|  |  |  || || ||
Db       7457 GGCGCGGCCGACCTGGGCGAGGACCTCGCGGGCCTCGGCGCGGAGGTGGCGTTCGCCGCC 7516

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
               |||| ||  | |||  ||    ||    |      ||  |||| | ||| |  | ||
Db       7517 GCCGACGCCGCCGACCGCGAGAGCCTGGCGCGGGCGATCGCCACCGTGCCCGCCGAGCAT 7576

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              || || ||  ||||| | ||||| || ||  || ||||||||||| ||  |        |||
Db       7577 CCGCTGACGGCCGTCGTGCACACGGCGGGAGTCGTCGACGACGCGACGGTGGAGGCGCTC 7636

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              || ||     |  ||  ||    || || ||| | || | | ||  |||       ||| |||
Db       7637 ACACCGGAACGGCTGGACGCGGTACTGCGCCCGAAGGTCGACGCCGCGTGGAACCTGCAC 7696

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
               | |||||| |  ||   |  |||  |||||||||||||| |||||||| | ||| || |
Db       7697 GAGCTCACCAAGGACCTGCGGCTCGACGCCTTCGTCCTCTTCTCCTCCGTCTCCGGCATC 7756

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              |||||  |  ||||||| |||||||||||| || |||||| |   ||||||||||||||||
Db       7757 GTCGGCACCGCCGGCCAGGCCAACTACGCGGCGGCCAACACGGGCCTCGACGCCCTCGCC 7816

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
               |||||||  | ||     |||  |||||||     | |||||||||  ||||
Db       7817 GCCCACCGCGCCGCCACGGGCCTGGCCGCCACGTCGCTGGCCTGGGGCCTCTGG 7870


Search completed: June 18, 2003, 05:08:24
Job time : 66.1981 secs
```

```
Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              |  |||      | |||| | |  |||||| || ||||||||| ||||||||| ||||||
Db       3763 TGGGGCAGCGGCAGCCAGGGCGTCTACGCGGCGGCCAACGCCCACCTCGACGCGCTCGCC 3822

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
                ||  |||| | |||    |    | || |||    ||||||||||  ||||
Db       3823 GCCCGGCGCCGCGCCCGGGGCGAGACGGCGACCTCGGTCGCCTGGGGCCTCTGG 3876


RESULT 13
US-09-861-289-34
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-34

  Query Match             40.1%;  Score 214;  DB 10;  Length 4689;
  Best Local Similarity   63.9%;  Pred. No. 9.1e-39;
  Matches  341;  Conservative    0;  Mismatches  190;  Indels    3;  Gaps    1;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||||||||||||||| ||||| ||||||| || ||||  | | | |     |
Db       3346 GGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGATG 3405

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
               ||  ||||  | ||    |||||||||||||||  |||||| | ||||   ||  |||||
Db       3406 GCCCACCACGGAGCCG---AACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCC 3462

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              ||    || |||||| || |||||      |   |||  || | |||||||| || ||
Db       3463 GGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCACCATCGCCGCC 3522

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||  |  | ||||  ||    | |      ||||||| || |||||||| |  |
Db       3523 TGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCCCGCCGAGACG 3582

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              ||||||||| ||||| |||||||||| |||   | | || ||   | ||   |  | |
Db       3583 CCCCTCACCGCCGTCGTCCACACCGCCGGCGCACCGGGCGGCGATCCGCTGGACGTCACC 3642

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              |||      |  ||  | | ||||  |||| ||   |    | |||||  |  ||||| ||
Db       3643 GGCCCGGAGGACATCGCCCGCATCCTGGGCGCGAAGACGAGCGGCGCCGAGGTCCTCGAC 3702

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | ||  ||    ||| || ||  |||||||||||||||||||||    |||||    |
Db       3703 GACCTGCTCCGCGGCACTCCGCTGGACGCCTTCGTCCTCTACTCCTCGAACGCCGGGGTC 3762

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              |  |||      | |||| | |  |||||| || ||||||||| ||||||||| ||||||
Db       3763 TGGGGCAGCGGCAGCCAGGGCGTCTACGCGGCGGCCAACGCCCACCTCGACGCGCTCGCC 3822

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
                ||  |||| | |||    |    | || |||    ||||||||||  ||||
Db       3823 GCCCGGCGCCGCGCCCGGGGCGAGACGGCGACCTCGGTCGCCTGGGGCCTCTGG 3876


RESULT 14
US-09-860-846-1
; Sequence 1, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-1

  Query Match             36.8%;  Score 196.4;  DB 9;  Length 15872;
  Best Local Similarity   60.5%;  Pred. No. 6.6e-35;
  Matches  323;  Conservative    0;  Mismatches  211;  Indels    0;  Gaps    0;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||| ||||||| || ||||| | |||||    | | |||| ||| |||||||
Db       7337 GGCACCGTCCTCGTCACCGGGGGCACCGGAGGCCTCGGAGCGCTCCTCGCCCGCCACCTC 7396

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
                 |  ||        |  |  || || ||    |||||   |||  || || |||||
Db       7397 GTGGAGCGTCACGGGGTGCGCCGGCTGCTGCTGGTGAGCCGCCGCGGGGCCGACGCCCCC 7456

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              |||      ||||    |     | |||          |||    |  |   |  || || ||
Db       7457 GGCGCGGCCGACCTGGGCGAGGACCTCGCGGGCCTCGGCGCGGAGGTGGCGTTCGCCGCC 7516

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCCACAACAC 240
              ||||  ||  | |||  ||     ||     |       ||  |||| | ||| |  | ||
Db       7517 GCCGACGCCGCCGACCGCGAGAGCCTGGCGCGGGCGATCGCCACCGTGCCCGCCGAGCAT 7576

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              || || ||  ||||| | ||||| || ||  || |||||||||| ||  |     |||
Db       7577 CCGCTGACGGCCGTCGTGCACACGGCGGGAGTCGTCGACGACGCGACGGTGGAGGCGCTC 7636

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              || ||     | ||  ||   || || ||| | || | | ||  |||       ||| |||
Db       7637 ACACCGGAACGGCTGGACGCGGTACTGCGCCCGAAGGTCGACGCCGCGTGGAACCTGCAC 7696

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | ||||||  |  ||   |  |||  |||||||||||||| |||||||  | ||| || |
Db       7697 GAGCTCACCAAGGACCTGCGGCTCGACGCCTTCGTCCTCTTCTCCTCCGTCTCCGGCATC 7756

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              |||||  |  ||||||| |||||||||| || ||||||  |   ||||||||||||||||
Db       7757 GTCGGCACCGCCGGCCAGGCCAACTACGCGGCGGCCAACACGGGCCTCGACGCCCTCGCC 7816

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              |||||||  | ||     |||  |||||||      | |||||||||  ||||
Db       7817 GCCCACCGCGCCGCCACGGGCCTGGCCGCCACGTCGCTGGCCTGGGGCCTCTGG 7870


RESULT 15
US-09-988-384B-1
; Sequence 1, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
```

```
Db       3463 GGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCACCATCGCCGCC 3522

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||  |  | ||||  ||    | |     ||||||  || ||||||||  |  |
Db       3523 TGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCCCGCCGAGACG 3582

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              |||||||||  ||||| ||||||||||  |||    |  | ||  ||    | ||    | |  |
Db       3583 CCCCTCACCGCCGTCGTCCACACCGCCGGCGCACCGGGCGGCGATCCGCTGGACGTCACC 3642

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
                |||      |  ||  |  | ||||  |||| ||  |   | |||||  |  ||||| ||
Db       3643 GGCCCGGAGGACATCGCCCGCATCCTGGGCGCGAAGACGAGCGGCGCCGAGGTCCTCGAC 3702

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | ||   ||    ||| || ||.  ||||||||||||||||||||    |||||    |
Db       3703 GACCTGCTCCGCGGCACTCCGCTGGACGCCTTCGTCCTCTACTCCTCGAACGCCGGGGTC 3762

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              |  |||      | |||| | |  |||||| || |||||||| ||||||||| ||||||
Db       3763 TGGGGCAGCGGCAGCCAGGGCGTCTACGCGGCGGCCAACGCCCACCTCGACGCGCTCGCC 3822

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
                ||  |||| | |||     |    | || |||     |||||||||||  ||||
Db       3823 GCCCGGCGCCGCGCCCGGGGCGAGACGGCGACCTCGGTCGCCTGGGGCCTCTGG 3876


RESULT 11
US-09-988-384B-34
; Sequence 34, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-34

  Query Match             40.1%;  Score 214;  DB 9;  Length 4689;
  Best Local Similarity   63.9%;  Pred. No. 9.1e-39;
  Matches  341;  Conservative    0;  Mismatches  190;  Indels    3;  Gaps    1;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||||||||||||||| ||||| ||||||| || ||||  | | | |    |
Db       3346 GGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGATG 3405

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
                ||  ||||  | ||    |||||||||||||  ||||||  | ||||   ||  |||||
Db       3406 GCCCACCACGGAGCCG---AACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCC 3462

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
                ||    || |||||| || |||||      |   |||  || | |||||||| || ||
Db       3463 GGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCACCATCGCCGCC 3522

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||  |  | ||||  ||    | |     ||||||  || ||||||||  |  |
Db       3523 TGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCCCGCCGAGACG 3582

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              |||||||||  ||||| ||||||||||  |||    |  | ||  ||    | ||    | |  |
Db       3583 CCCCTCACCGCCGTCGTCCACACCGCCGGCGCACCGGGCGGCGATCCGCTGGACGTCACC 3642

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
                |||      |  ||  |  | ||||  |||| ||  |   | |||||  |  ||||| ||
Db       3643 GGCCCGGAGGACATCGCCCGCATCCTGGGCGCGAAGACGAGCGGCGCCGAGGTCCTCGAC 3702

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | ||   ||    ||| || ||.  ||||||||||||||||||||    |||||    |
Db       3703 GACCTGCTCCGCGGCACTCCGCTGGACGCCTTCGTCCTCTACTCCTCGAACGCCGGGGTC 3762

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              |  |||      | |||| | |  |||||| || |||||||| ||||||||| ||||||
Db       3763 TGGGGCAGCGGCAGCCAGGGCGTCTACGCGGCGGCCAACGCCCACCTCGACGCGCTCGCC 3822

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
                ||  |||| | |||     |    | || |||     |||||||||||  ||||
Db       3823 GCCCGGCGCCGCGCCCGGGGCGAGACGGCGACCTCGGTCGCCTGGGGCCTCTGG 3876


RESULT 12
US-09-836-821-34
; Sequence 34, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-34

  Query Match             40.1%;  Score 214;  DB 9;  Length 4689;
  Best Local Similarity   63.9%;  Pred. No. 9.1e-39;
  Matches  341;  Conservative    0;  Mismatches  190;  Indels    3;  Gaps    1;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||||||||||||||| ||||| ||||||| || ||||  | | | |    |
Db       3346 GGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGATG 3405

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
                ||  ||||  | ||    |||||||||||||  ||||||  | ||||   ||  |||||
Db       3406 GCCCACCACGGAGCCG---AACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCC 3462

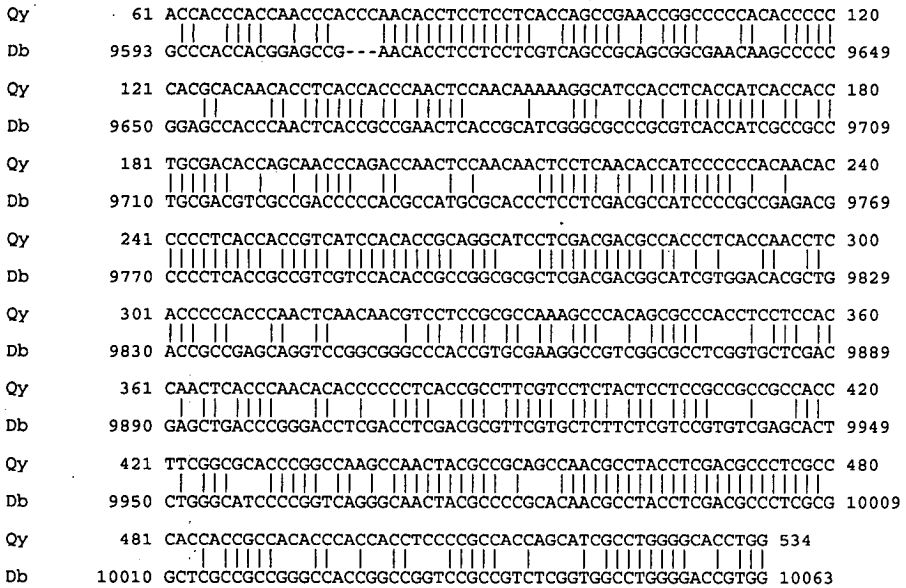
Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
                ||    || |||||| || |||||      |   |||  || | |||||||| || ||
Db       3463 GGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCACCATCGCCGCC 3522

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||  |  | ||||  ||    | |     ||||||  || ||||||||  |  |
Db       3523 TGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCCCGCCGAGACG 3582

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              |||||||||  ||||| ||||||||||  |||    |  | ||  ||    | ||    | |  |
Db       3583 CCCCTCACCGCCGTCGTCCACACCGCCGGCGCACCGGGCGGCGATCCGCTGGACGTCACC 3642

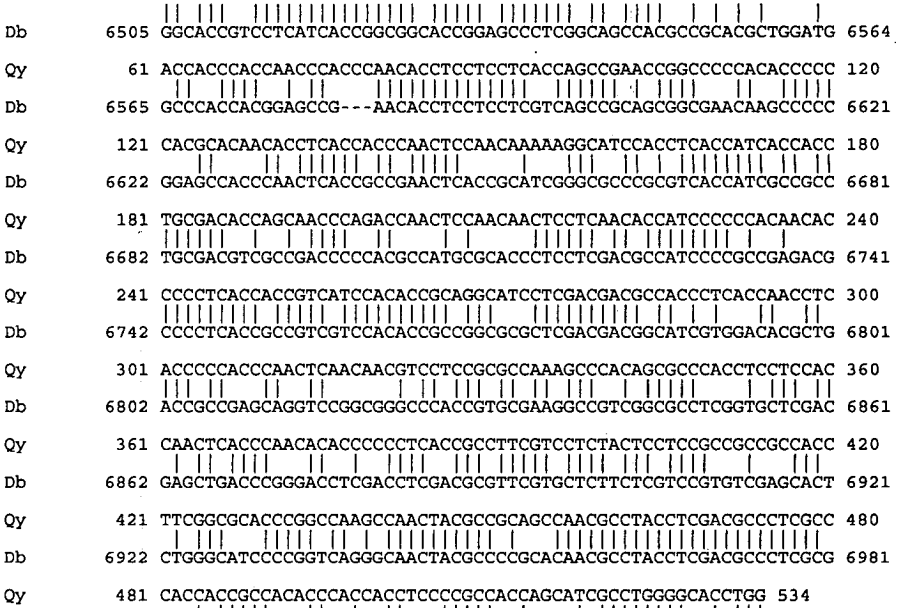
Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
                |||      |  ||  |  | ||||  |||| ||  |   | |||||  |  ||||| ||
Db       3643 GGCCCGGAGGACATCGCCCGCATCCTGGGCGCGAAGACGAGCGGCGCCGAGGTCCTCGAC 3702

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | ||   ||    ||| || ||.  ||||||||||||||||||||    |||||    |
Db       3703 GACCTGCTCCGCGGCACTCCGCTGGACGCCTTCGTCCTCTACTCCTCGAACGCCGGGGTC 3762
```



RESULT 9
US-09-793-708-19
; Sequence 19, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002121
; CURRENT APPLICATION NUMBER: US/09/793,708
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/846,247
; PRIOR FILING DATE: 1997-04-30
; PRIOR APPLICATION NUMBER: US 60/134,990
; PRIOR FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-793-708-19

Query Match 41.9%; Score 223.6; DB 9; Length 38506;
Best Local Similarity 65.0%; Pred. No. 6.6e-41;
Matches 347; Conservative 0; Mismatches 184; Indels 3; Gaps 1;



RESULT 10
US-09-860-846-34
; Sequence 34, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
; LENGTH: 4689
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-34

Query Match 40.1%; Score 214; DB 9; Length 4689;
Best Local Similarity 63.9%; Pred. No. 9.1e-39;
Matches 341; Conservative 0; Mismatches 190; Indels 3; Gaps 1;

```
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-5

  Query Match             41.9%;  Score 223.6;  DB 9;  Length 36778;
  Best Local Similarity   65.0%;  Pred. No. 6.6e-41;
  Matches  347;  Conservative    0;  Mismatches  184;  Indels    3;  Gaps    1;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||||||||||||||| ||||| ||||||| || ||||  | | | |    |
Db       8363 GGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGATG 8422

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
              ||  ||||  | ||    |||||||||||||||  ||||||  | ||||   ||  ||||||
Db       8423 GCCCACCACGGAGCCG---AACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCC 8479

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              ||    || ||||||  || |||||      |   |||  ||  | ||||||||  || ||
Db       8480 GGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCACCATCGCCGCC 8539

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||  |  | ||||  ||    | |     ||||||  || ||||||||  |  |
Db       8540 TGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCCCGCCGAGACG 8599

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              |||||||||  |||||  ||||||||||  |||    ||||||||||  ||  |  |    ||  ||
Db       8600 CCCCTCACCGCCGTCGTCCACACCGCCGGCGCGCTCGACGACGGCATCGTGGACACGCTG 8659

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              |||  ||   ||  ||      | || |||  ||  ||  |||   | |||||     | |||  ||
Db       8660 ACCGCCGAGCAGGTCCGGCGGGCCCACCGTGCGAAGGCCGTCGGCGCCTCGGTGCTCGAC 8719

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | || ||||   ||  |  ||||  ||| |||||  |||| |||  ||||   |   |||
Db       8720 GAGCTGACCCGGGACCTCGACCTCGACGCGTTCGTGCTCTTCTCGTCCGTGTCGAGCACT 8779

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              | |||   |||||  || | ||||||||||  |   ||||||||||||||||||||||||
Db       8780 CTGGGCATCCCCGGTCAGGGCAACTACGCCCCGCACAACGCCTACCTCGACGCCCTCGCG 8839

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              | |||||   ||  |  ||   |||||  |    | ||||||||  | |||
Db       8840 GCTCGCCGCCGGGCCACCGGCCGGTCCGCCGTCTCGGTGGCCTGGGGACCGTGG 8893


RESULT 7
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-5

  Query Match             41.9%;  Score 223.6;  DB 10;  Length 36778;
  Best Local Similarity   65.0%;  Pred. No. 6.6e-41;
  Matches  347;  Conservative    0;  Mismatches  184;  Indels    3;  Gaps    1;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||||||||||||||| ||||| ||||||| || ||||  | | | |    |
Db       8363 GGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGATG 8422

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
              ||  ||||  | ||    |||||||||||||||  ||||||  | ||||   ||  ||||||
Db       8423 GCCCACCACGGAGCCG---AACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCC 8479

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              ||    || ||||||  || |||||      |   |||  ||  | ||||||||  || ||
Db       8480 GGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCACCATCGCCGCC 8539

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||  |  | ||||  ||    | |     ||||||  || ||||||||  |  |
Db       8540 TGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCCCGCCGAGACG 8599

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              |||||||||  |||||  ||||||||||  |||    ||||||||||  ||  |  |    ||  ||
Db       8600 CCCCTCACCGCCGTCGTCCACACCGCCGGCGCGCTCGACGACGGCATCGTGGACACGCTG 8659

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              |||  ||   ||  ||      | || |||  ||  ||  |||   | |||||     | |||  ||
Db       8660 ACCGCCGAGCAGGTCCGGCGGGCCCACCGTGCGAAGGCCGTCGGCGCCTCGGTGCTCGAC 8719

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | || ||||   ||  |  ||||  ||| |||||  |||| |||  ||||   |   |||
Db       8720 GAGCTGACCCGGGACCTCGACCTCGACGCGTTCGTGCTCTTCTCGTCCGTGTCGAGCACT 8779

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              | |||   |||||  || | ||||||||||  |   ||||||||||||||||||||||||
Db       8780 CTGGGCATCCCCGGTCAGGGCAACTACGCCCCGCACAACGCCTACCTCGACGCCCTCGCG 8839

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              | |||||   ||  |  ||   |||||  |    | ||||||||  | |||
Db       8840 GCTCGCCGCCGGGCCACCGGCCGGTCCGCCGTCTCGGTGGCCTGGGGACCGTGG 8893


RESULT 8
US-09-988-384B-5
; Sequence 5, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 5
;   LENGTH: 37948
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-5

  Query Match             41.9%;  Score 223.6;  DB 9;  Length 37948;
  Best Local Similarity   65.0%;  Pred. No. 6.6e-41;
  Matches  347;  Conservative    0;  Mismatches  184;  Indels    3;  Gaps    1;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||||||||||||||| ||||| ||||||| || ||||  | | | |    |
Db       9533 GGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGATG 9592
```

```
Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              | |||||   ||  |  ||   |||||  |    | ||||||||  | |||
Db       7099 GCTCGCCGCCGGGCCACCGGCCGGTCCGCCGTCTCGGTGGCCTGGGGACCGTGG 7152


RESULT 4
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-30

  Query Match             41.9%;  Score 223.6;  DB 10;  Length 13842;
  Best Local Similarity   65.0%;  Pred. No. 6.8e-41;
  Matches  347;  Conservative    0;  Mismatches  184;  Indels    3;  Gaps    1;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||||||||||||||| ||||| |||||| || ||||  | | | |    |
Db       6622 GGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGATG 6681

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
               ||  ||||  | ||    |||||||||||||  ||||||  | ||||   ||  |||||
Db       6682 GCCCACCACGGAGCCG---AACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCC 6738

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
                 ||    || |||||| || |||||      |   |||  || | |||||||| || ||
Db       6739 GGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCACCATCGCCGCC 6798

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||  |  | ||||  ||    | |     |||||| || |||||||| | |
Db       6799 TGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCCCGCCGAGACG 6858

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              |||||||||| ||||| |||||||||| |||   |||||||||| || | |   ||  ||
Db       6859 CCCCTCACCGCCGTCGTCCACACCGCCGGCGCGCTCGACGACGGCATCGTGGACACGCTG 6918

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              ||| ||   ||  ||       | || ||| || || |||  | |||||    | ||| ||
Db       6919 ACCGCCGAGCAGGTCCGGCGGGCCCACCGTGCGAAGGCCGTCGGCGCCTCGGTGCTCGAC 6978

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | || ||||   ||  |  ||||  ||| ||||| |||| ||| ||||   |   |||
Db       6979 GAGCTGACCCGGGACCTCGACCTCGACGCGTTCGTGCTCTTCTCGTCCGTGTCGAGCACT 7038

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              | |||   ||||| || | ||||||||||  |   ||||||||||||||||||||||
Db       7039 CTGGGCATCCCCGGTCAGGGCAACTACGCCCCGCACAACGCCTACCTCGACGCCCTCGCG 7098

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              | |||||   ||  |  ||   |||||  |    | ||||||||  | |||
Db       7099 GCTCGCCGCCGGGCCACCGGCCGGTCCGCCGTCTCGGTGGCCTGGGGACCGTGG 7152


RESULT 5
US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-5

  Query Match             41.9%;  Score 223.6;  DB 9;  Length 36778;
  Best Local Similarity   65.0%;  Pred. No. 6.6e-41;
  Matches  347;  Conservative    0;  Mismatches  184;  Indels    3;  Gaps    1;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||||||||||||||| ||||| |||||| || ||||  | | | |    |
Db       8363 GGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGATG 8422

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
               ||  ||||  | ||    |||||||||||||  ||||||  | ||||   ||  |||||
Db       8423 GCCCACCACGGAGCCG---AACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCC 8479

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
                 ||    || |||||| || |||||      |   |||  || | |||||||| || ||
Db       8480 GGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCACCATCGCCGCC 8539

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||  |  | ||||  ||    | |     |||||| || |||||||| | |
Db       8540 TGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCCCGCCGAGACG 8599

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              |||||||||| ||||| |||||||||| |||   |||||||||| || | |   ||  ||
Db       8600 CCCCTCACCGCCGTCGTCCACACCGCCGGCGCGCTCGACGACGGCATCGTGGACACGCTG 8659

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              ||| ||   ||  ||       | || ||| || || |||  | |||||    | ||| ||
Db       8660 ACCGCCGAGCAGGTCCGGCGGGCCCACCGTGCGAAGGCCGTCGGCGCCTCGGTGCTCGAC 8719

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | || ||||   ||  |  ||||  ||| ||||| |||| ||| ||||   |   |||
Db       8720 GAGCTGACCCGGGACCTCGACCTCGACGCGTTCGTGCTCTTCTCGTCCGTGTCGAGCACT 8779

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              | |||   ||||| || | ||||||||||  |   ||||||||||||||||||||||
Db       8780 CTGGGCATCCCCGGTCAGGGCAACTACGCCCCGCACAACGCCTACCTCGACGCCCTCGCG 8839

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              | |||||   ||  |  ||   |||||  |    | ||||||||  | |||
Db       8840 GCTCGCCGCCGGGCCACCGGCCGGTCCGCCGTCTCGGTGGCCTGGGGACCGTGG 8893


RESULT 6
US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
```

```
Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              ||||||||| ||||| |||||||||| |||    |||||||||| || | |    ||  ||
Db       6859 CCCCTCACCGCCGTCGTCCACACCGCCGGCGCGCTCGACGACGGCATCGTGGACACGCTG 6918

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              ||| ||   ||  ||       | || ||| || || |||  | |||||     | ||| ||
Db       6919 ACCGCCGAGCAGGTCCGGCGGGCCCACCGTGCGAAGGCCGTCGGCGCCTCGGTGCTCGAC 6978

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | || ||||   ||  |  ||||  ||| ||||| |||| ||| ||||   |   |||
Db       6979 GAGCTGACCCGGGACCTCGACCTCGACGCGTTCGTGCTCTTCTCGTCCGTGTCGAGCACT 7038

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              | |||   ||||| || | |||||||||| |   ||||||||||||||||||||||||
Db       7039 CTGGGCATCCCCGGTCAGGGCAACTACGCCCCGCACAACGCCTACCTCGACGCCCTCGCG 7098

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              | |||||   ||  |  ||   |||||  |    | ||||||||  | |||
Db       7099 GCTCGCCGCCGGGCCACCGGCCGGTCCGCCGTCTCGGTGGCCTGGGGACCGTGG 7152


RESULT 2
US-09-988-384B-30
; Sequence 30, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-30

  Query Match             41.9%;  Score 223.6;  DB 9;  Length 13842;
  Best Local Similarity   65.0%;  Pred. No. 6.8e-41;
  Matches  347;  Conservative    0;  Mismatches  184;  Indels    3;  Gaps    1;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||||||||||||| ||||| ||||||| || ||||  | | | |    |
Db       6622 GGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGATG 6681

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
              ||  ||||  | ||    |||||||||||||  |||||| | ||||   ||  |||||
Db       6682 GCCCACCACGGAGCCG---AACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCC 6738

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
                ||    || |||||| || |||||      |   |||  || | |||||||| || ||
Db       6739 GGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCACCATCGCCGCC 6798

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||  |  | ||||  ||    | |     |||||| || |||||||| |  |
Db       6799 TGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCCCGCCGAGACG 6858

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              ||||||||| ||||| |||||||||| |||   |||||||||| || | |   ||  ||
Db       6859 CCCCTCACCGCCGTCGTCCACACCGCCGGCGCGCTCGACGACGGCATCGTGGACACGCTG 6918

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              ||| ||   ||  ||       | || ||| || || |||  | |||||     | ||| ||
Db       6919 ACCGCCGAGCAGGTCCGGCGGGCCCACCGTGCGAAGGCCGTCGGCGCCTCGGTGCTCGAC 6978

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | || ||||   ||  |  ||||  ||| ||||| |||| ||| ||||   |   |||
Db       6979 GAGCTGACCCGGGACCTCGACCTCGACGCGTTCGTGCTCTTCTCGTCCGTGTCGAGCACT 7038

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              | |||   ||||| || | |||||||||| |   ||||||||||||||||||||||||
Db       7039 CTGGGCATCCCCGGTCAGGGCAACTACGCCCCGCACAACGCCTACCTCGACGCCCTCGCG 7098

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              | |||||   ||  |  ||   |||||  |    | ||||||||  | |||
Db       7099 GCTCGCCGCCGGGCCACCGGCCGGTCCGCCGTCTCGGTGGCCTGGGGACCGTGG 7152


RESULT 3
US-09-836-821-30
; Sequence 30, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-30

  Query Match             41.9%;  Score 223.6;  DB 9;  Length 13842;
  Best Local Similarity   65.0%;  Pred. No. 6.8e-41;
  Matches  347;  Conservative    0;  Mismatches  184;  Indels    3;  Gaps    1;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||||||||||||| ||||| ||||||| || ||||  | | | |    |
Db       6622 GGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGATG 6681

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
              ||  ||||  | ||    |||||||||||||  |||||| | ||||   ||  |||||
Db       6682 GCCCACCACGGAGCCG---AACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCC 6738

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
                ||    || |||||| || |||||      |   |||  || | |||||||| || ||
Db       6739 GGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCACCATCGCCGCC 6798

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||  |  | ||||  ||    | |     |||||| || |||||||| |  |
Db       6799 TGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCCCGCCGAGACG 6858

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              ||||||||| ||||| |||||||||| |||   |||||||||| || | |   ||  ||
Db       6859 CCCCTCACCGCCGTCGTCCACACCGCCGGCGCGCTCGACGACGGCATCGTGGACACGCTG 6918

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              ||| ||   ||  ||       | || ||| || || |||  | |||||     | ||| ||
Db       6919 ACCGCCGAGCAGGTCCGGCGGGCCCACCGTGCGAAGGCCGTCGGCGCCTCGGTGCTCGAC 6978

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | || ||||   ||  |  ||||  ||| ||||| |||| ||| ||||   |   |||
Db       6979 GAGCTGACCCGGGACCTCGACCTCGACGCGTTCGTGCTCTTCTCGTCCGTGTCGAGCACT 7038

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              | |||   ||||| || | |||||||||| |   ||||||||||||||||||||||||
Db       7039 CTGGGCATCCCCGGTCAGGGCAACTACGCCCCGCACAACGCCTACCTCGACGCCCTCGCG 7098
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 17:56:34 ; Search time 63.1072 Seconds
(without alignments)
12253.192 Million cell updates/sec

Title: US-09-914-286-1_COPY_5143_5676
Perfect score: 534
Sequence: 1 ggaaccaccctcatcaccgg.....gcatcgcctggggcacctgg 534

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	223.6	41.9	13842	9	US-09-860-846-30	Sequence 30, Appl
	2	223.6	41.9	13842	9	US-09-988-384B-30	Sequence 30, Appl
	3	223.6	41.9	13842	9	US-09-836-821-30	Sequence 30, Appl
	4	223.6	41.9	13842	10	US-09-861-289-30	Sequence 30, Appl
	5	223.6	41.9	36778	9	US-09-860-846-5	Sequence 5, Appli
	6	223.6	41.9	36778	9	US-09-836-821-5	Sequence 5, Appli
	7	223.6	41.9	36778	10	US-09-861-289-5	Sequence 5, Appli
	8	223.6	41.9	37948	9	US-09-988-384B-5	Sequence 5, Appli
	9	223.6	41.9	38506	9	US-09-793-708-19	Sequence 19, Appl
	10	214	40.1	4689	9	US-09-860-846-34	Sequence 34, Appl
	11	214	40.1	4689	9	US-09-988-384B-34	Sequence 34, Appl
	12	214	40.1	4689	9	US-09-836-821-34	Sequence 34, Appl
	13	214	40.1	4689	10	US-09-861-289-34	Sequence 34, Appl
	14	196.4	36.8	15872	9	US-09-860-846-1	Sequence 1, Appli
	15	196.4	36.8	15872	9	US-09-988-384B-1	Sequence 1, Appli
	16	196.4	36.8	15872	9	US-09-836-821-1	Sequence 1, Appli
	17	196.4	36.8	15872	10	US-09-861-289-1	Sequence 1, Appli
	18	185.2	34.7	11220	9	US-09-860-846-32	Sequence 32, Appl
	19	185.2	34.7	11220	9	US-09-988-384B-32	Sequence 32, Appl
	20	185.2	34.7	11220	9	US-09-836-821-32	Sequence 32, Appl
	21	185.2	34.7	11220	10	US-09-861-289-32	Sequence 32, Appl
c	22	156.2	29.3	12733	9	US-10-032-393-47	Sequence 47, Appl
c	23	156.2	29.3	12739	9	US-10-032-393-8	Sequence 8, Appli
	24	154.6	29.0	50937	9	US-09-808-880-1	Sequence 1, Appli
	25	119.2	22.3	925	9	US-09-735-056-1	Sequence 1, Appli
c	26	114.6	21.5	1075	10	US-09-864-761-19241	Sequence 19241, A
c	27	114.6	21.5	1403	10	US-09-864-761-2513	Sequence 2513, Ap
	28	110.4	20.7	552	9	US-09-957-483-32	Sequence 32, Appl
	29	110.4	20.7	68750	9	US-10-014-717-1	Sequence 1, Appli
	30	109.2	20.4	504	10	US-09-878-574-4330	Sequence 4330, Ap
	31	109.2	20.4	1064	10	US-09-804-682-29	Sequence 29, Appl
c	32	107.6	20.1	446	10	US-09-864-761-20699	Sequence 20699, A
	33	107.2	20.1	552	9	US-09-957-483-34	Sequence 34, Appl
	34	106.8	20.0	1065	10	US-09-804-682-33	Sequence 33, Appl
	35	104.4	19.6	1635	10	US-09-864-761-20241	Sequence 20241, A
	36	104.4	19.6	1973	10	US-09-864-761-3471	Sequence 3471, Ap
	37	103	19.3	15720	9	US-10-025-380-1058	Sequence 1058, Ap
	38	103	19.3	15720	10	US-09-922-217-1058	Sequence 1058, Ap
	39	103	19.3	15720	10	US-09-833-263-1058	Sequence 1058, Ap
c	40	99.8	18.7	155074	9	US-10-026-188-6	Sequence 6, Appli
c	41	98.6	18.5	931	9	US-10-198-846-8585	Sequence 8585, Ap
	42	93	17.4	53522	9	US-09-904-968A-1	Sequence 1, Appli
c	43	92.4	17.3	615	9	US-10-125-815-3	Sequence 3, Appli
c	44	91.2	17.1	1173	9	US-09-894-844-60	Sequence 60, Appl
c	45	90.4	16.9	6306	9	US-10-239-676-223	Sequence 223, App

ALIGNMENTS

RESULT 1
US-09-860-846-30
; Sequence 30, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-30

Query Match 41.9%; Score 223.6; DB 9; Length 13842;
Best Local Similarity 65.0%; Pred. No. 6.8e-41;
Matches 347; Conservative 0; Mismatches 184; Indels 3; Gaps 1;

```
Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||||||||||||||| ||||| ||||||| || ||||  | | | |    |
Db       6622 GGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGATG 6681

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
                ||  ||||  | ||    |||||||||||||||  |||||| | ||||   ||  |||||
Db       6682 GCCCACCACGGAGCCG---AACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCC 6738

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
                 ||    || |||||| || |||||      |   |||  || | |||||||| || ||
Db       6739 GGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCACCATCGCCGCC 6798

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCCACAACAC 240
              |||||| |  | ||||  ||    | |      |||||| || |||||||| | |
Db       6799 TGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCCCGCCGAGACG 6858
```

```
Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCCACACCCCC 120
Db      13516 GTCGCGAAACACAGCGCCAAACACCTGCTCCTCACCTCGAGGAAGGGCGCGCGTGCTCCG 13575

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
Db      13576 GGCGCGGAGGCTCTGCGAAGCGAGCTCGAAGCGCTGGGGGCCTCGGTCACCCTCGTCGCG 13635

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
Db      13636 TGCGACGTGGCCGACCCACGCGCCCTCCGGACCCTCCTGGACAGCATCCCGAGGGATCAT 13695

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
Db      13696 CCGATCACGGCCGTCGTGCACGCCGCCGGCGCCCTCGACGACGGGCCGCTCGGTAGCATG 13755

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
Db      13756 AGCGCCGAGCGCATCGCTCGCGTCTTTGACCCCAAGCTCGATGCCGCTTGGTACTTGCAT 13815

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
Db      13816 GAGCTCACCCAGGACGAGCCGGTCGCGGCCTTCGTCCTCTTCTCGGCCGCCTCCGGCGTC 13875

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
Db      13876 CTTGGTGGTCCAGGTCAGTCGAACTACGCCGCTGCCAATGCCTTCCTCGATGCGCTCGCA 13935

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
Db      13936 CATCACCGGCGCGCCCAAGGACTCCCAGCCGCTTCGCTCGCCTGGGGCTACTGG 13989


RESULT 15
US-08-729-214-6
; Sequence 6, Application US/08729214
; Patent No. 5817502
;  GENERAL INFORMATION:
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  van Pee, Karl-Heinz
;    APPLICANT:  Kirner, Sabine
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  520 White Plains Road
;      CITY:  Tarrytown
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10591
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/729,214
;      FILING DATE:  TBA
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-729-214-6

  Query Match           36.8%;  Score 196.4;  DB 1;  Length 28958;
  Best Local Similarity   60.5%;  Pred. No. 2.3e-28;
  Matches  323;  Conservative    0;  Mismatches  211;  Indels    0;  Gaps    0;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
Db      13456 GGGACCGTCCTCATCACGGGAGGCACCGGGACGCTAGGAGTCCTGGTCGCACGCCACCTC 13515

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
Db      13516 GTCGCGAAACACAGCGCCAAACACCTGCTCCTCACCTCGAGGAAGGGCGCGCGTGCTCCG 13575

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
Db      13576 GGCGCGGAGGCTCTGCGAAGCGAGCTCGAAGCGCTGGGGGCCTCGGTCACCCTCGTCGCG 13635

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
Db      13636 TGCGACGTGGCCGACCCACGCGCCCTCCGGACCCTCCTGGACAGCATCCCGAGGGATCAT 13695

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
Db      13696 CCGATCACGGCCGTCGTGCACGCCGCCGGCGCCCTCGACGACGGGCCGCTCGGTAGCATG 13755

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
Db      13756 AGCGCCGAGCGCATCGCTCGCGTCTTTGACCCCAAGCTCGATGCCGCTTGGTACTTGCAT 13815

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
Db      13816 GAGCTCACCCAGGACGAGCCGGTCGCGGCCTTCGTCCTCTTCTCGGCCGCCTCCGGCGTC 13875

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
Db      13876 CTTGGTGGTCCAGGTCAGTCGAACTACGCCGCTGCCAATGCCTTCCTCGATGCGCTCGCA 13935

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
Db      13936 CATCACCGGCGCGCCCAAGGACTCCCAGCCGCTTCGCTCGCCTGGGGCTACTGG 13989


Search completed: June 18, 2003, 01:06:58
Job time : 27.1323 secs
```

```
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Ciba-Geigy Corporation
;       STREET:  520 White Plains Road, P.O. Box 2005
;       CITY:  Tarrytown
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10591
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/764,233A
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/729,214
;       FILING DATE:  09-OCT-1996
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/258,261
;       FILING DATE:  08-JUN-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Meigs, J. Timothy
;       REGISTRATION NUMBER:  38,241
;       REFERENCE/DOCKET NUMBER:  1506/CIP6
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (919) 541-8587
;       TELEFAX:  (919) 541-8689
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  28958 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     ORIGINAL SOURCE:
;       ORGANISM:  Sorangium cellulosum
;     IMMEDIATE SOURCE:
;       CLONE:  p98/1
US-08-764-233A-4

  Query Match           36.8%;  Score 196.4;  DB 1;  Length 28958;
  Best Local Similarity   60.5%;  Pred. No. 2.3e-28;
  Matches  323;  Conservative    0;  Mismatches  211;  Indels    0;  Gaps    0;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  ||||||||| || || |||||  | || |    ||   || | | |||||||
Db      13456 GGGACCGTCCTCATCACGGGAGGCACCGGGACGCTAGGAGTCCTGGTCGCACGCCACCTC 13515

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
              | |  | ||   | || ||||||| |||||||||    | |  ||| | |   | ||
Db      13516 GTCGCGAAACACAGCGCCAAACACCTGCTCCTCACCTCGAGGAAGGGCGCGCGTGCTCCG 13575

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              |||  |    ||    | | | ||| ||       ||    |    ||||| ||  | |
Db      13576 GGCGCGGAGGCTCTGCGAAGCGAGCTCGAAGCGCTGGGGGCCTCGGTCACCCTCGTCGCG 13635

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||     | |||||  |   ||||     |||||  ||| ||||||     | ||
Db      13636 TGCGACGTGGCCGACCCACGCGCCCTCCGGACCCTCCTGGACAGCATCCCGAGGGATCAT 13695

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              ||  ||||  ||||| | ||| |||| |||  ||||||||||   | |||   | | |
Db      13696 CCGATCACGGCCGTCGTGCACGCCGCCGGCGCCCTCGACGACGGGCCGCTCGGTAGCATG 13755

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              | | ||   |   ||     |||| |   | ||||   | |   |||      | | ||
Db      13756 AGCGCCGAGCGCATCGCTCGCGTCTTTGACCCCAAGCTCGATGCCGCTTGGTACTTGCAT 13815

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | ||||||||  ||   ||  || | ||||||||||||||| |||  ||||| ||| | |
Db      13816 GAGCTCACCCAGGACGAGCCGGTCGCGGCCTTCGTCCTCTTCTCGGCCGCCTCCGGCGTC 13875

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              | ||| |  || || ||  | |||||||||| ||||| |||| |||||| || |||||
Db      13876 CTTGGTGGTCCAGGTCAGTCGAACTACGCCGCTGCCAATGCCTTCCTCGATGCGCTCGCA 13935

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              || ||||| | | ||||    ||||| ||| |     |||||||||| ||||
Db      13936 CATCACCGGCGCGCCCAAGGACTCCCAGCCGCTTCGCTCGCCTGGGGCTACTGG 13989


RESULT 14
US-08-457-335A-6
; Sequence 6, Application US/08457335A
; Patent No. 5723759
;  GENERAL INFORMATION:
;     APPLICANT:  Schupp, Thomas
;     APPLICANT:  Ligon, James M.
;     APPLICANT:  Beck, James Joseph
;     APPLICANT:  Hill, Dwight Steven
;     APPLICANT:  Ryals, John Andrew
;     APPLICANT:  Gaffney, Thomas Deane
;     APPLICANT:  Lam, Stephen Ting
;     APPLICANT:  Hammer, Phillip E.
;     APPLICANT:  Uknes, Scott Joseph
;     TITLE OF INVENTION:  Genes for the synthesis of
;     TITLE OF INVENTION:  antipathogenic substances
;     NUMBER OF SEQUENCES:  22
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Ciba-Geigy Corporation
;       STREET:  7 Skyline Drive
;       CITY:  Hawthorne
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10532
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/457,335A
;       FILING DATE:  01-JUN-1995
;       CLASSIFICATION:  800
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/457,205
;       FILING DATE:  01-JUN-1995
;       APPLICATION NUMBER:  08/258,261
;       FILING DATE:  08-Jun-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Elmer, James Scott
;       REGISTRATION NUMBER:  36,129
;       REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  919-541-8614
;       TELEFAX:  919-541-8689
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  28958 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
US-08-457-335A-6

  Query Match           36.8%;  Score 196.4;  DB 1;  Length 28958;
  Best Local Similarity   60.5%;  Pred. No. 2.3e-28;
  Matches  323;  Conservative    0;  Mismatches  211;  Indels    0;  Gaps    0;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  ||||||||| || || |||||  | || |    ||   || | | |||||||
Db      13456 GGGACCGTCCTCATCACGGGAGGCACCGGGACGCTAGGAGTCCTGGTCGCACGCCACCTC 13515
```

```
Db      13456 GGGACCGTCCTCATCACGGGAGGCACCGGGACGCTAGGAGTCCTGGTCGCACGCCACCTC 13515

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120

Db      13516 GTCGCGAAACACAGCGCCAAACACCTGCTCCTCACCTCGAGGAAGGGCGCGCGTGCTCCG 13575

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180

Db      13576 GGCGCGGAGGCTCTGCGAAGCGAGCTCGAAGCGCTGGGGGCCTCGGTCACCCTCGTCGCG 13635

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240

Db      13636 TGCGACGTGGCCGACCCACGCGCCCTCCGGACCCTCCTGGACAGCATCCCGAGGGATCAT 13695

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300

Db      13696 CCGATCACGGCCGTCGTGCACGCCGCCGGCGCCCTCGACGACGGGCCGCTCGGTAGCATG 13755

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360

Db      13756 AGCGCCGAGCGCATCGCTCGCGTCTTTGACCCCAAGCTCGATGCCGCTTGGTACTTGCAT 13815

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420

Db      13816 GAGCTCACCCAGGACGAGCCGGTCGCGGCCTTCGTCCTCTTCTCGGCCGCCTCCGGCGTC 13875

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480

Db      13876 CTTGGTGGTCCAGGTCAGTCGAACTACGCCGCTGCCAATGCCTTCCTCGATGCGCTCGCA 13935

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534

Db      13936 CATCACCGGCGCGCCCAAGGACTCCCAGCCGCTTCGCTCGCCTGGGGCTACTGG 13989
```

RESULT 12
US-08-458-076A-6
; Sequence 6, Application US/08458076A
; Patent No. 5698425
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,076A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-458-076A-6

Query Match 36.8%; Score 196.4; DB 1; Length 28958;
Best Local Similarity 60.5%; Pred. No. 2.3e-28;
Matches 323; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

```
Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60

Db      13456 GGGACCGTCCTCATCACGGGAGGCACCGGGACGCTAGGAGTCCTGGTCGCACGCCACCTC 13515

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120

Db      13516 GTCGCGAAACACAGCGCCAAACACCTGCTCCTCACCTCGAGGAAGGGCGCGCGTGCTCCG 13575

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180

Db      13576 GGCGCGGAGGCTCTGCGAAGCGAGCTCGAAGCGCTGGGGGCCTCGGTCACCCTCGTCGCG 13635

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240

Db      13636 TGCGACGTGGCCGACCCACGCGCCCTCCGGACCCTCCTGGACAGCATCCCGAGGGATCAT 13695

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300

Db      13696 CCGATCACGGCCGTCGTGCACGCCGCCGGCGCCCTCGACGACGGGCCGCTCGGTAGCATG 13755

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360

Db      13756 AGCGCCGAGCGCATCGCTCGCGTCTTTGACCCCAAGCTCGATGCCGCTTGGTACTTGCAT 13815

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420

Db      13816 GAGCTCACCCAGGACGAGCCGGTCGCGGCCTTCGTCCTCTTCTCGGCCGCCTCCGGCGTC 13875

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480

Db      13876 CTTGGTGGTCCAGGTCAGTCGAACTACGCCGCTGCCAATGCCTTCCTCGATGCGCTCGCA 13935

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534

Db      13936 CATCACCGGCGCGCCCAAGGACTCCCAGCCGCTTCGCTCGCCTGGGGCTACTGG 13989
```

RESULT 13
US-08-764-233A-4
; Sequence 4, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neff, Snezanna
; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
; NUMBER OF SEQUENCES: 10

```
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,342
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-342-6

  Query Match             36.8%;  Score 196.4;  DB 1;  Length 28958;
  Best Local Similarity   60.5%;  Pred. No. 2.3e-28;
  Matches  323;  Conservative    0;  Mismatches  211;  Indels    0;  Gaps    0;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||||||||| || || |||||  | || |     ||   || | | |||||||
Db      13456 GGGACCGTCCTCATCACGGGAGGCACCGGGACGCTAGGAGTCCTGGTCGCACGCCACCTC 13515

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
                | |  | ||   | || ||||||| |||||||||     | |  ||| | |    | ||
Db      13516 GTCGCGAAACACAGCGCCAAACACCTGCTCCTCACCTCGAGGAAGGGCGCGCGTGCTCCG 13575

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
                |||  |    ||    | | | ||| ||      ||   |    ||||| ||  | |
Db      13576 GGCGCGGAGGCTCTGCGAAGCGAGCTCGAAGCGCTGGGGGCCTCGGTCACCCTCGTCGCG 13635

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||     | |||||  |   ||||     |||||  ||| ||||||     | ||
Db      13636 TGCGACGTGGCCGACCCACGCGCCCTCCGGACCCTCCTGGACAGCATCCCGAGGGATCAT 13695

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              ||  ||||  ||||| | ||| |||| |||  |||||||||||    | |||   | | |
Db      13696 CCGATCACGGCCGTCGTGCACGCCGCCGGCGCCCTCGACGACGGGCCGCTCGGTAGCATG 13755

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              | | ||  |   ||     |||| |   | ||||   | |   |||     | | ||
Db      13756 AGCGCCGAGCGCATCGCTCGCGTCTTTGACCCCAAGCTCGATGCCGCTTGGTACTTGCAT 13815

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
                | |||||||  ||   ||  || | |||||||||||||| |||  ||||| ||| |  |
Db      13816 GAGCTCACCCAGGACGAGCCGGTCGCGGCCTTCGTCCTCTTCTCGGCCGCCTCCGGCGTC 13875

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              | || |  || || ||  | ||||||||||| ||||| |||| |||||| || |||||
Db      13876 CTTGGTGGTCCAGGTCAGTCGAACTACGCCGCTGCCAATGCCTTCCTCGATGCGCTCGCA 13935

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              || ||||| | | ||||    ||||| ||| |     ||||||||||  ||||
Db      13936 CATCACCGGCGCGCCCAAGGACTCCCAGCCGCTTCGCTCGCCTGGGGCTACTGG 13989


RESULT 11
US-08-457-646A-6
; Sequence 6, Application US/08457646A
; Patent No. 5679560
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,646A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-646A-6

  Query Match             36.8%;  Score 196.4;  DB 1;  Length 28958;
  Best Local Similarity   60.5%;  Pred. No. 2.3e-28;
  Matches  323;  Conservative    0;  Mismatches  211;  Indels    0;  Gaps    0;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
```

```
Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||||||||| || || |||||  | || |     ||   || | | |||||||
Db      13456 GGGACCGTCCTCATCACGGGAGGCACCGGGACGCTAGGAGTCCTGGTCGCACGCCACCTC 13515

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
                | |  | ||   | || |||||||  |||||||||     | |  ||| | |   | ||
Db      13516 GTCGCGAAACACAGCGCCAAACACCTGCTCCTCACCTCGAGGAAGGGCGCGCGTGCTCCG 13575

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              |||  |   ||   | | | ||| ||      ||   |  |||||  ||  | |
Db      13576 GGCGCGGAGGCTCTGCGAAGCGAGCTCGAAGCGCTGGGGGCCTCGGTCACCCTCGTCGCG 13635

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||     | |||||  |   ||||     |||||  ||| ||||||     | ||
Db      13636 TGCGACGTGGCCGACCCACGCGCCCTCCGGACCCTCCTGGACAGCATCCCGAGGGATCAT 13695

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              ||  ||||  |||||  | |||  |||| |||  ||||||||||     | |||   | | |
Db      13696 CCGATCACGGCCGTCGTGCACGCCGCCGGCGCCCTCGACGACGGGCCGCTCGGTAGCATG 13755

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              | | ||  |   ||     |||| |   | ||||   | |   |||      | | ||
Db      13756 AGCGCCGAGCGCATCGCTCGCGTCTTTGACCCCAAGCTCGATGCCGCTTGGTACTTGCAT 13815

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | ||||||||  ||   ||  || | ||||||||||||| |||  ||||| ||| |  |
Db      13816 GAGCTCACCCAGGACGAGCCGGTCGCGGCCTTCGTCCTCTTCTCGGCCGCCTCCGGCGTC 13875

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              | || |  || || ||  | |||||||||| ||||| |||| |||||| || |||||
Db      13876 CTTGGTGGTCCAGGTCAGTCGAACTACGCCGCTGCCAATGCCTTCCTCGATGCGCTCGCA 13935

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              || ||||| | | ||||    |||||  ||| |     ||||||||||  ||||
Db      13936 CATCACCGGCGCGCCCAAGGACTCCCAGCCGCTTCGCTCGCCTGGGGCTACTGG 13989
```

RESULT 9
US-08-456-837-6
; Sequence 6, Application US/08456837
; Patent No. 5643774
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,837
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-456-837-6

Query Match 36.8%; Score 196.4; DB 1; Length 28958;
Best Local Similarity 60.5%; Pred. No. 2.3e-28;
Matches 323; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

```
Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||||||||| || || |||||  | || |     ||   || | | |||||||
Db      13456 GGGACCGTCCTCATCACGGGAGGCACCGGGACGCTAGGAGTCCTGGTCGCACGCCACCTC 13515

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
                | |  | ||   | || |||||||  |||||||||     | |  ||| | |   | ||
Db      13516 GTCGCGAAACACAGCGCCAAACACCTGCTCCTCACCTCGAGGAAGGGCGCGCGTGCTCCG 13575

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              |||  |   ||   | | | ||| ||      ||   |  |||||  ||  | |
Db      13576 GGCGCGGAGGCTCTGCGAAGCGAGCTCGAAGCGCTGGGGGCCTCGGTCACCCTCGTCGCG 13635

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||     | |||||  |   ||||     |||||  ||| ||||||     | ||
Db      13636 TGCGACGTGGCCGACCCACGCGCCCTCCGGACCCTCCTGGACAGCATCCCGAGGGATCAT 13695

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              ||  ||||  |||||  | |||  |||| |||  ||||||||||     | |||   | | |
Db      13696 CCGATCACGGCCGTCGTGCACGCCGCCGGCGCCCTCGACGACGGGCCGCTCGGTAGCATG 13755

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              | | ||  |   ||     |||| |   | ||||   | |   |||      | | ||
Db      13756 AGCGCCGAGCGCATCGCTCGCGTCTTTGACCCCAAGCTCGATGCCGCTTGGTACTTGCAT 13815

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | ||||||||  ||   ||  || | ||||||||||||| |||  ||||| ||| |  |
Db      13816 GAGCTCACCCAGGACGAGCCGGTCGCGGCCTTCGTCCTCTTCTCGGCCGCCTCCGGCGTC 13875

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              | || |  || || ||  | |||||||||| ||||| |||| |||||| || |||||
Db      13876 CTTGGTGGTCCAGGTCAGTCGAACTACGCCGCTGCCAATGCCTTCCTCGATGCGCTCGCA 13935

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              || ||||| | | ||||    |||||  ||| |     ||||||||||  ||||
Db      13936 CATCACCGGCGCGCCCAAGGACTCCCAGCCGCTTCGCTCGCCTGGGGCTACTGG 13989
```

RESULT 10
US-08-457-342-6
; Sequence 6, Application US/08457342
; Patent No. 5662898
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting

```
Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              |||||| |   | |||||  |   ||||||   ||| |  || |||||||       |||
Db      12513 TGCGACGCGGCCGACCCACGCGCCCTCCAAGCCCTCTTGGACTCCATCCCGAGCGCTCAC 12572

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              || |||||| ||||| ||||| |||| |||  |||||||||| |   |||  |  | |
Db      12573 CCGCTCACCGCCGTCGTCCACGCCGCGGGCGCCCTCGACGACGGCCTGCTCGGCGCCATG 12632

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              | ||||   |   || ||  |||| |   | ||||   | |     ||    | | | ||
Db      12633 AGCCCCGAGCGCATCGACCGCGTCTTTGCCCCCAAGCTCGATGCTGCTTGGCACTTGCAT 12692

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | |||||||| |||  |||||| |||||||||||||| ||| ||||| || | |  |
Db      12693 GAGCTCACCCAAGACAAGCCCCTCGCCGCCTTCGTCCTCTTCTCGTCCGCTGCTGGCGTC 12752

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              | ||    || || ||  | |||||||||| |||| |||| |||||| || |||||
Db      12753 CTTGGTAGTCCAGGTCAGTCGAACTACGCCGCGGCCAATGCCTTCCTCGATGCGCTCGCG 12812

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              || ||||| |   |||||   |||| ||| ||    |||| ||||||   |||
Db      12813 CATCACCGGCGTGCCCACGGGCTCCCGGCCTCCTCGCTCGCATGGGGCTATTGG 12866
```

RESULT 7
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-1

Query Match 36.8%; Score 196.4; DB 4; Length 15872;
Best Local Similarity 60.5%; Pred. No. 2.2e-28;
Matches 323; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

```
Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||| ||||||| || ||||| | |||||   | | |||| ||| |||||||
Db       7337 GGCACCGTCCTCGTCACCGGGGGCACCGGAGGCCTCGGAGCGCTCCTCGCCCGCCACCTC 7396

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
                 |  ||       | |  || || ||    |||||   |||  || || |||||
Db       7397 GTGGAGCGTCACGGGGTGCGCCGGCTGCTGCTGGTGAGCCGCCGCGGGGCCGACGCCCCC 7456

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
               |||     ||||  |   | | |||      |||    |  |  |  || || ||
Db       7457 GGCGCGGCCGACCTGGGCGAGGACCTCGCGGGCCTCGGCGCGGAGGTGGCGTTCGCCGCC 7516

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              |||| ||  | |||  ||    ||     |       ||  |||| | |||  |  | ||
Db       7517 GCCGACGCCGCCGACCGCGAGAGCCTGGCGCGGGCGATCGCCACCGTGCCCGCCGAGCAT 7576

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              || || ||  ||||| | ||||| || ||  || ||||||||| ||  |       |||
Db       7577 CCGCTGACGGCCGTCGTGCACACGGCGGGAGTCGTCGACGACGCGACGGTGGAGGCGCTC 7636

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              || ||    |  ||  ||   || || ||| | || | | ||  |||      ||| |||
```

```
Db       7637 ACACCGGAACGGCTGGACGCGGTACTGCGCCCGAAGGTCGACGCCGCGTGGAACCTGCAC 7696

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | |||||| |  ||    |  |||  |||||||||||||| |||||||| | ||| || |
Db       7697 GAGCTCACCAAGGACCTGCGGCTCGACGCCTTCGTCCTCTTCTCCTCCGTCTCCGGCATC 7756

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              ||||| |  |||||||| |||||||||| || |||||| |   ||||||||||||||||
Db       7757 GTCGGCACCGCCGGCCAGGCCAACTACGCGGCGGCCAACACGGGCCTCGACGCCCTCGCC 7816

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              |||||||  | ||     |||  |||||||     | |||||||||  ||||
Db       7817 GCCCACCGCGCCGCCACGGGCCTGGCCGCCACGTCGCTGGCCTGGGGCCTCTGG 7870
```

RESULT 8
US-08-258-261B-6
; Sequence 6, Application US/08258261B
; Patent No. 5639949
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,261B
; FILING DATE: 08-JUN-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-258-261B-6

Query Match 36.8%; Score 196.4; DB 1; Length 28958;
Best Local Similarity 60.5%; Pred. No. 2.3e-28;
Matches 323; Conservative 0; Mismatches 211; Indels 0; Gaps 0;

```
Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              || ||    | |||     |||||||||| | || ||| ||  ||  | |   || |||
Db      18890 ACACCGCAGCGGCTCGCGGCCGTCCTCCGCCCGAAGGCCGACGCCGTACGCAATCTGCAC 18949

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              ||||||||||       | ||||  | || ||| |||||||||| || |||||||
Db      18950 GAACTCACCCAGGGGCACGCCCTGTCGGCGTTCATCCTCTACTCGTCGGCCGCCGGAGTG 19009

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              |||||    | |||||   |  ||||||||| |||||||||||||| ||| || |||||
Db      19010 CTCGGCAGCGCGGGGCCAGAGCGGCTACGCCGCCGCCAACGCCTACCTGGACTCCTTCGCC 19069

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
                |   || |  | ||     || ||||||       ||| ||||||| | |||
Db      19070 GTCTGGCGGCGGAGCCGGGGACTGCCCGCCGTATCGCTCGGCTGGGGCCCGTGG 19123
```

RESULT 6
US-08-764-233A-1
; Sequence 1, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neff, Snezanna
; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,233A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: 1506/CIP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Sorangium cellulosum
; IMMEDIATE SOURCE:
; CLONE: p98/1, pJL3, and pVKM15
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 383..760
; OTHER INFORMATION: /product= "SorR"
; OTHER INFORMATION: /note= "This gene encodes a protein that is highly homologous
; OTHER INFORMATION: the reductase domains of type I PKSs such as eryA from
; OTHER INFORMATION: Saccharopolyspora erythraea."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 927..19874
; OTHER INFORMATION: /product= "SorA"
; OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKSs that
; OTHER INFORMATION: are known to be involved in the synthesis of polyketide
; OTHER INFORMATION: compounds."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 942..7115
; OTHER INFORMATION: /product= "Module 1 of SorA"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7203..12884
; OTHER INFORMATION: /product= "Module 2 of SorA"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 13455..19616
; OTHER INFORMATION: /product= "Module 3 of SorA"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19871..46318
; OTHER INFORMATION: /product= "SorB"
; OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKS genes
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19870..24556
; OTHER INFORMATION: /product= "Module 1 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 24638..30820
; OTHER INFORMATION: /product= "Module 2 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30881..35446
; OTHER INFORMATION: /product= "Module 3 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 35528..40114
; OTHER INFORMATION: /product= "Module 4 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 40190..46318
; OTHER INFORMATION: /product= "Module 5 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 46851..47891
; OTHER INFORMATION: /product= "SorM"
; OTHER INFORMATION: /note= "The protein encoded by the sorM gene is highly
; OTHER INFORMATION: homologous to the methyltransferase from Streptomyces
; OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the
; OTHER INFORMATION: polyketide rappamicin."
US-08-764-233A-1

Query Match 38.9%; Score 207.6; DB 1; Length 49377;
Best Local Similarity 61.8%; Pred. No. 2.1e-30;
Matches 330; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

```
Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  ||||||||| || || |||||| | || |   |||   || | | ||| |||
Db      12333 GGCACCGTCCTCATCACGGGAGGCACCGGCACGCTAGGAGCCCTGGTCGCACGCCATCTC 12392

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
              | |    ||   | || | ||||| || |||||      |    ||| | ||| | ||
Db      12393 GTCGCAAGACACGGCGCCAAGCACCTGCTTCTCACGTCGAGGCAGGGCGCGCACGCTCCG 12452

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              |||  |   |       ||| | ||| ||       ||   |   |||| |   | |
Db      12453 GGCGCCGAGGCCTCGCGAACCGAGCTCGAAGCGCTGGGGGCCTCTGTCACACTTCGCGCG 12512
```

```
Db       6982 GCTCGCCGCCGGGCCACCGGCCGGTCCGCCGTCTCGGTGGCCTGGGGACCGTGG 7035


RESULT 4
US-09-105-537-34
; Sequence 34, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-34

  Query Match             40.1%;  Score 214;  DB 4;  Length 4689;
  Best Local Similarity   63.9%;  Pred. No. 1.2e-31;
  Matches  341;  Conservative    0;  Mismatches  190;  Indels    3;  Gaps    1;

Qy           1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
Db        3346 GGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGATG 3405

Qy          61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
Db        3406 GCCCACCACGGAGCCG---AACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCC 3462

Qy         121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
Db        3463 GGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCACCATCGCCGCC 3522

Qy         181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
Db        3523 TGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCCCGCCGAGACG 3582

Qy         241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
Db        3583 CCCCTCACCGCCGTCGTCCACACCGCCGGCGCACCGGGCGGCGATCCGCTGGACGTCACC 3642

Qy         301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
Db        3643 GGCCCGGAGGACATCGCCCGCATCCTGGGCGCGAAGACGAGCGGCGCCGAGGTCCTCGAC 3702

Qy         361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
Db        3703 GACCTGCTCCGCGGCACTCCGCTGGACGCCTTCGTCCTCTACTCCTCGAACGCCGGGGTC 3762

Qy         421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
Db        3763 TGGGGCAGCGGCAGCCAGGGCGTCTACGCGGCGGCCAACGCCCACCTCGACGCGCTCGCC 3822

Qy         481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
Db        3823 GCCCGGCGCCGCGCCCGGGGCGAGACGGCGACCTCGGTCGCCTGGGGCCTCTGG 3876


RESULT 5
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  43280 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  816..14234
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14351..19945
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20010..31199
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31232..36067
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36249..41774
US-08-804-227C-1

  Query Match             39.5%;  Score 210.8;  DB 2;  Length 43280;
  Best Local Similarity   62.2%;  Pred. No. 5.4e-31;
  Matches  332;  Conservative    0;  Mismatches  202;  Indels    0;  Gaps    0;

Qy           1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
Db       18590 GGCACCGTCCTGATCACTGGGGGCACCGGCAGCCTCGGCGCGGAGGCGGCCCGCCATCTG 18649

Qy          61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
Db       18650 GTCACCCGGCACGGTGCCCGGCGCCTGCTCCTGACCAGCCGAAGCGGCCCGCAGGCCCCC 18709

Qy         121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
Db       18710 GGCGCGGCCGAACTCGTCGCCGAACTGGCCGCCTTGGGCGCCCACGCGGACGTGGCCGCC 18769

Qy         181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
Db       18770 TGCGACGTCGCCGACCGCGCCGCCCTGCGGGCCCTGCTCGACCGCGTACCCGCCGGCCAC 18829

Qy         241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
Db       18830 CCGCTGACCGCGGTCCTGCACACGGCGGGCGTCCTGGACGACGGCGTCCTCACCGCCCAG 18889
```

```
Db       6979 GAGCTGACCCGGGACCTCGACCTCGACGCGTTCGTGCTCTTCTCGTCCGTGTCGAGCACT 7038

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480

Db       7039 CTGGGCATCCCCGGTCAGGGCAACTACGCCCCGCACAACGCCTACCTCGACGCCCTCGCG 7098

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534

Db       7099 GCTCGCCGCCGGGCCACCGGCCGGTCCGCCGTCTCGGTGGCCTGGGGACCGTGG 7152


RESULT 2
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-5

  Query Match           41.9%;  Score 223.6;  DB 4;  Length 36778;
  Best Local Similarity   65.0%;  Pred. No. 2.4e-33;
  Matches  347;  Conservative    0;  Mismatches  184;  Indels    3;  Gaps    1;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60

Db       8363 GGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGATG 8422

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120

Db       8423 GCCCACCACGGAGCCG---AACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCC 8479

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180

Db       8480 GGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCACCATCGCCGCC 8539

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240

Db       8540 TGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCCCGCCGAGACG 8599

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300

Db       8600 CCCCTCACCGCCGTCGTCCACACCGCCGGCGCGCTCGACGACGGCATCGTGGACACGCTG 8659

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360

Db       8660 ACCGCCGAGCAGGTCCGGCGGGCCCACCGTGCGAAGGCCGTCGGCGCCTCGGTGCTCGAC 8719

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420

Db       8720 GAGCTGACCCGGGACCTCGACCTCGACGCGTTCGTGCTCTTCTCGTCCGTGTCGAGCACT 8779

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480

Db       8780 CTGGGCATCCCCGGTCAGGGCAACTACGCCCCGCACAACGCCTACCTCGACGCCCTCGCG 8839

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534

Db       8840 GCTCGCCGCCGGGCCACCGGCCGGTCCGCCGTCTCGGTGGCCTGGGGACCGTGG 8893


RESULT 3
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-19

  Query Match           41.9%;  Score 223.6;  DB 3;  Length 38506;
  Best Local Similarity   65.0%;  Pred. No. 2.4e-33;
  Matches  347;  Conservative    0;  Mismatches  184;  Indels    3;  Gaps    1;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60

Db       6505 GGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGATG 6564

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120

Db       6565 GCCCACCACGGAGCCG---AACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCC 6621

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180

Db       6622 GGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCACCATCGCCGCC 6681

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240

Db       6682 TGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCCCGCCGAGACG 6741

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300

Db       6742 CCCCTCACCGCCGTCGTCCACACCGCCGGCGCGCTCGACGACGGCATCGTGGACACGCTG 6801

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360

Db       6802 ACCGCCGAGCAGGTCCGGCGGGCCCACCGTGCGAAGGCCGTCGGCGCCTCGGTGCTCGAC 6861

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420

Db       6862 GAGCTGACCCGGGACCTCGACCTCGACGCGTTCGTGCTCTTCTCGTCCGTGTCGAGCACT 6921

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480

Db       6922 CTGGGCATCCCCGGTCAGGGCAACTACGCCCCGCACAACGCCTACCTCGACGCCCTCGCG 6981

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 23.1323 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_5143_5676
Perfect score: 534
Sequence: 1 ggaaccaccctcatcaccgg.....gcatcgcctggggcacctgg 534

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	223.6	41.9	13842	4	US-09-105-537-30	Sequence 30, Appl
2	223.6	41.9	36778	4	US-09-105-537-5	Sequence 5, Appli
3	223.6	41.9	38506	3	US-09-320-878-19	Sequence 19, Appl
4	214	40.1	4689	4	US-09-105-537-34	Sequence 34, Appl
5	210.8	39.5	43280	2	US-08-804-227C-1	Sequence 1, Appli
6	207.6	38.9	49377	1	US-08-764-233A-1	Sequence 1, Appli
7	196.4	36.8	15872	4	US-09-105-537-1	Sequence 1, Appli
8	196.4	36.8	28958	1	US-08-258-261B-6	Sequence 6, Appli
9	196.4	36.8	28958	1	US-08-456-837-6	Sequence 6, Appli
10	196.4	36.8	28958	1	US-08-457-342-6	Sequence 6, Appli
11	196.4	36.8	28958	1	US-08-457-646A-6	Sequence 6, Appli
12	196.4	36.8	28958	1	US-08-458-076A-6	Sequence 6, Appli
13	196.4	36.8	28958	1	US-08-764-233A-4	Sequence 4, Appli
14	196.4	36.8	28958	1	US-08-457-335A-6	Sequence 6, Appli
15	196.4	36.8	28958	1	US-08-729-214-6	Sequence 6, Appli
16	196.4	36.8	28958	3	US-09-028-934-6	Sequence 6, Appli
17	191.6	35.9	44377	2	US-08-804-227C-7	Sequence 7, Appli
18	191.6	35.9	44377	2	US-08-804-198-1	Sequence 1, Appli
19	185.2	34.7	11220	4	US-09-105-537-32	Sequence 32, Appl
20	180.4	33.8	13987	2	US-08-804-227C-13	Sequence 13, Appl
21	177.2	33.2	11219	1	US-07-642-734C-1	Sequence 1, Appli
22	177.2	33.2	11219	3	US-08-439-009A-1	Sequence 1, Appli
23	175.2	32.8	477	3	US-09-010-809-2	Sequence 2, Appli
c 24	175.2	32.8	477	3	US-09-010-809-17	Sequence 17, Appl
25	154.6	29.0	50937	4	US-09-428-517-1	Sequence 1, Appli
26	129.2	24.2	80161	3	US-09-036-987A-1	Sequence 1, Appli
27	129.2	24.2	80161	4	US-09-370-700-1	Sequence 1, Appli
28	128	24.0	20235	1	US-07-642-734C-3	Sequence 3, Appli
29	128	24.0	20235	3	US-08-439-009A-3	Sequence 3, Appli
30	124.2	23.3	33529	4	US-09-144-085-3	Sequence 3, Appli
31	119.2	22.3	925	3	US-08-858-003-1	Sequence 1, Appli
32	119.2	22.3	925	3	US-09-078-166-1	Sequence 1, Appli
33	119.2	22.3	925	4	US-08-997-467-1	Sequence 1, Appli
34	117.4	22.0	897	4	US-09-434-288-6	Sequence 6, Appli
35	117.4	22.0	1926	4	US-09-249-585A-4	Sequence 4, Appli
36	117.4	22.0	1931	2	US-09-130-114-2	Sequence 2, Appli
37	110.4	20.7	68750	3	US-09-335-409-1	Sequence 1, Appli
38	110.4	20.7	68750	4	US-09-568-102-1	Sequence 1, Appli
39	110.4	20.7	68750	4	US-09-567-969-1	Sequence 1, Appli
40	110.4	20.7	68750	4	US-09-568-480-1	Sequence 1, Appli
41	110.4	20.7	68750	4	US-09-568-486-1	Sequence 1, Appli
42	110.4	20.7	68750	4	US-09-568-472-1	Sequence 1, Appli
43	110.4	20.7	68750	4	US-09-567-899-1	Sequence 1, Appli
44	110.4	20.7	71989	4	US-09-443-501A-2	Sequence 2, Appli
c 45	110	20.6	4403765	4	US-09-103-840A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-30

Query Match 41.9%; Score 223.6; DB 4; Length 13842;
Best Local Similarity 65.0%; Pred. No. 2.3e-33;
Matches 347; Conservative 0; Mismatches 184; Indels 3; Gaps 1;

```
Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||||||||||||| ||||| ||||||| || ||||  | | | |    |
Db       6622 GGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCACGCCGCACGCTGGATG 6681

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
                ||  ||||  | ||    |||||||||||||  |||||| | ||||   ||  |||||
Db       6682 GCCCACCACGGAGCCG---AACACCTCCTCCTCGTCAGCCGCAGCGGCGAACAAGCCCCC 6738

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
               ||    || |||||| || |||||      |   |||  || | |||||||| || ||
Db       6739 GGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGCGTCACCATCGCCGCC 6798

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||  |  | ||||  ||    | |      |||||| || |||||||| |  |
Db       6799 TGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCCATCCCCGCCGAGACG 6858

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              ||||||||| ||||| |||||||||| |||   |||||||||| || | |   ||  ||
Db       6859 CCCCTCACCGCCGTCGTCCACACCGCCGGCGCGCTCGACGACGGCATCGTGGACACGCTG 6918

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              ||| ||   ||  ||       | || ||| || || |||  | |||||    | ||| ||
Db       6919 ACCGCCGAGCAGGTCCGGCGGGCCCACCGTGCGAAGGCCGTCGGCGCCTCGGTGCTCGAC 6978

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
```

THIS PAGE BLANK (USPTO)

CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC peripheral neuropathies. They can also be used in agricultural and
CC veterinary applications.
XX
SQ Sequence 4571 BP; 659 A; 1829 C; 1469 G; 614 T; 0 other;

Query Match 45.7%; Score 244; DB 21; Length 4571;
Best Local Similarity 67.7%; Pred. No. 3.1e-31;
Matches 359; Conservative 0; Mismatches 165; Indels 6; Gaps 1;

```
Qy          5 CCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCA 64
              ||  || || ||||  |  | ||  | ||  | ||  | ||| |   |||  | ||| ||
Db       3943 CCCTCCACACCACCACCCCACCCACCACCACCCCCCTCAACCCCGAACACGCCATCATCA 4002

Qy         65 CCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACG 124
               |  |     | ||     |||| |  || |  |     |  ||   ||  |||||||
Db       4003 TCACCGGCGGCTCCGGCACCCTCGCCGGCATCCTCGCCCGCCACCTGAACCACCCCCACA 4062

Qy        125 CACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCG 184
              |  | | |||| ||  |   |  |  |   | | || || |  ||||  |  | ||||||
Db       4063 CCTACCTCCTCTCCCGCACCCCACCCCCCGACGCCACCCCCGGCACCCACCTCCCCTGCG 4122

Qy        185 ACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCC 244
              ||  | ||  ||||  |||||||.|         |||||  |  |||||||       || ||||
Db       4123 ACGTCGGCGACCCCCACCAACTCGCCACCACCCTCACCCACATCCCC------CAACCCC 4176

Qy        245 TCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCC 304
              |||||  ||  ||  |||||||||| | || ||||||||||| ||  ||||   |  ||||||||
Db       4177 TCACCGCCATCTTCCACACCGCCGCCACCCTCGACGACGGCATCCTCCACGCCCTCACCC 4236

Qy        305 CCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAAC 364
              ||  ||  |||| || |||||||| | |||||||| ||  ||||   | ||| ||||| |
Db       4237 CCGACCGCCTCACCACCGTCCTCCACCCCAAAGCCAACGCCGCCTGGCACCTGCACCACC 4296

Qy        365 TCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCG 424
              ||||||||  ||   |||||||||  |||||||||||||||  ||||||||||  | |||
Db       4297 TCACCCAAAACCAACCCCTCACCCACTTCGTCCTCTACTCCAGCGCCGCCGCCGTCCTCG 4356

Qy        425 GCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACC 484
              ||   ||||| ||||  ||||||||||| |||||||||| ||||||||||||||||  ||
Db       4357 GCAGCCCCGGACAAGGAAACTACGCCGCCGCCAACGCCTTCCTCGACGCCCTCGCCACCC 4416

Qy        485 ACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              ||||||||||||  |  ||  |||||||||  ||||||||||||||  |||
Db       4417 ACCGCCACACCCTCGGCCAACCCGCCACCTCCATCGCCTGGGGCATGTGG 4466
```

Search completed: June 17, 2003, 13:48:28
Job time : 109.318 secs

```
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Reeves C,  Chu D,  Khosla C,  Santi D,  Wu K;
XX
DR   WPI; 2000-317716/27.
DR   P-PSDB; AAY84728.
XX
PT   New isolated polyketide synthase nucleic acid and polyketide compounds,
PT   useful for treating e.g. transplant rejection, uveitis, multiple
PT   sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT   peripheral neuropathy  -
XX
PS   Example 1; Page 85-88; 126pp; English.
XX
CC   The present sequence represents an NheII-XhoII fragment that encodes
CC   module 8 of te FK-520 polyketide synthase (PKS) gene cluster with the
CC   endogenous acyltransferase (AT) domain replaced by the AT domain of
CC   module 13 of the rapamycin PKS (which is specific for methylmalonyl CoA).
CC   FK-506 is a potent immunosuppressants, and acts through intial formation
CC   of an intermediate complex with protein immunophilins known as FK-506
CC   binding proteins. The nucleic acids are used for producing polyketide
CC   compounds. The polyketide compounds can be used as immunosuppressants to
CC   prevent or treat transplant rejection, graft-versus-host disease or
CC   uveitis. They can also be used for treating e.g. alopecia universalis,
CC   autoimmune chronic active hepatitis, inflammatory bowel disease,
CC   multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
CC   also have neurotrophic activity and can be used to promote neurite
CC   outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC   and in intact animals, they promote regrowth of damaged facial and
CC   sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC   the brain. They can also be used for treating e.g. Parkinson's disease,
CC   Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC   peripheral neuropathies. They can also be used in agricultural and
CC   veterinary applications.
XX
SQ   Sequence 4547 BP; 658 A; 1851 C; 1437 G; 601 T; 0 other;

  Query Match             45.7%;  Score 244;  DB 21;  Length 4547;
  Best Local Similarity   67.7%;  Pred. No. 3.1e-31;
  Matches  359;  Conservative    0;  Mismatches  165;  Indels    6;  Gaps    1;

Qy          5 CCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCA 64
              ||  || || ||||  |  | ||  | ||  | ||  | ||| |   |||  | ||| ||
Db       3919 CCCTCCACACCACCACCCCACCCACCACCACCCCCCTCAACCCCGAACACGCCATCATCA 3978

Qy         65 CCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACG 124
              |  |     | ||     ||||  |  || |  |     |  ||   ||  |||||||
Db       3979 TCACCGGCGGCTCCGGCACCCTCGCCGGCATCCTCGCCCGCCACCTGAACCACCCCCACA 4038

Qy        125 CACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCG 184
              |  | | |||| ||  |   |  |  |   | | || || |  ||||  |  | ||||||
Db       4039 CCTACCTCCTCTCCCGCACCCCACCCCCCGACGCCACCCCCGGCACCCACCTCCCCTGCG 4098

Qy        185 ACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCC 244
              ||  | || ||||  |||||||       |||||  |  |||||||      || ||||
Db       4099 ACGTCGGCGACCCCCACCAACTCGCCACCACCCTCACCCACATCCCC------CAACCCC 4152

Qy        245 TCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCC 304
              ||||| || || |||||||||| | || |||||||||||| || ||||  |  ||||||||
Db       4153 TCACCGCCATCTTCCACACCGCCGCCACCCTCGACGACGGCATCCTCCACGCCCTCACCC 4212

Qy        305 CCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAAC 364
              ||  ||  |||| || |||||||| | ||||||| ||  ||||   | ||| ||||| |
Db       4213 CCGACCGCCTCACCACCGTCCTCCACCCCAAAGCCAACGCCGCCTGGCACCTGCACCACC 4272

Qy        365 TCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCG 424
              |||||||  ||   |||||||||  |||||||||||||||  |||||||||  | |||
Db       4273 TCACCCAAAACCAACCCCTCACCCACTTCGTCCTCTACTCCAGCGCCGCCGCCGTCCTCG 4332

Qy        425 GCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACC 484
              ||   ||||| ||||  ||||||||||| |||||||||| ||||||||||||||||  ||
Db       4333 GCAGCCCCGGACAAGGAAACTACGCCGCCGCCAACGCCTTCCTCGACGCCCTCGCCACCC 4392

Qy         485 ACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
               |||||||||||| |  ||  |||||||||  |||||||||||||||  |||
Db        4393 ACCGCCACACCCTCGGCCAACCCGCCACCTCCATCGCCTGGGGCATGTGG 4442


RESULT 15
AAA14662
ID   AAA14662 standard; DNA; 4571 BP.
XX
AC   AAA14662;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Nucleotide sequence of a fragment of the FK-520 PKS gene cluster.
XX
KW   FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW   immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW   transplant rejection; graft-versus-host disease; alopecia universalis;
KW   autoimmune chronic active hepatitis; inflammatory bowel disease;
KW   multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW   neurite outgrowth; nerve regrowth; Parkinson's disease;
KW   Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW   peripheral neuropathy; ss.
XX
OS   Synthetic.
OS   Streptomyces hygroscopicus.
XX
FH   Key             Location/Qualifiers
FT   CDS             9..4559
FT                   /*tag= a
FT                   /note= "no termination codon given"
XX
PN   WO200020601-A2.
XX
PD   13-APR-2000.
XX
PF   01-OCT-1999;   99WO-US22886.
XX
PR   02-OCT-1998;   98US-0102748.
PR   11-MAR-1999;   99US-0123810.
PR   17-JUN-1999;   99US-0139650.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Reeves C,  Chu D,  Khosla C,  Santi D,  Wu K;
XX
DR   WPI; 2000-317716/27.
DR   P-PSDB; AAY84726.
XX
PT   New isolated polyketide synthase nucleic acid and polyketide compounds,
PT   useful for treating e.g. transplant rejection, uveitis, multiple
PT   sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT   peripheral neuropathy  -
XX
PS   Example 1; Page 79-82; 126pp; English.
XX
CC   The present sequence represents an AvrII-XhoII fragment that encodes
CC   module 8 of te FK-520 polyketide synthase (PKS) gene cluster with the
CC   endogenous acyltransferase (AT) domain replaced by the AT domain of
CC   module 13 of the rapamycin PKS (which is specific for methylmalonyl CoA).
CC   FK-506 is a potent immunosuppressants, and acts through intial formation
CC   of an intermediate complex with protein immunophilins known as FK-506
CC   binding proteins. The nucleic acids are used for producing polyketide
CC   compounds. The polyketide compounds can be used as immunosuppressants to
CC   prevent or treat transplant rejection, graft-versus-host disease or
CC   uveitis. They can also be used for treating e.g. alopecia universalis,
CC   autoimmune chronic active hepatitis, inflammatory bowel disease,
CC   multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
CC   also have neurotrophic activity and can be used to promote neurite
CC   outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC   and in intact animals, they promote regrowth of damaged facial and
CC   sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC   the brain. They can also be used for treating e.g. Parkinson's disease,
```

```
DT   08-AUG-2000  (first entry)
XX
DE   Nucleotide sequence of a fragment of the FK-520 PKS gene cluster.
XX
KW   FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW   immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW   transplant rejection; graft-versus-host disease; alopecia universalis;
KW   autoimmune chronic active hepatitis; inflammatory bowel disease;
KW   multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW   neurite outgrowth; nerve regrowth; Parkinson's disease;
KW   Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW   peripheral neuropathy; ss.
XX
OS   Synthetic.
OS   Streptomyces hygroscopicus.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..4466
FT                   /*tag=  a
FT                   /note= "no termination codon given"
XX
PN   WO200020601-A2.
XX
PD   13-APR-2000.
XX
PF   01-OCT-1999;   99WO-US22886.
XX
PR   02-OCT-1998;   98US-0102748.
PR   11-MAR-1999;   99US-0123810.
PR   17-JUN-1999;   99US-0139650.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Reeves C,  Chu D,  Khosla C,  Santi D,  Wu K;
XX
DR   WPI; 2000-317716/27.
DR   P-PSDB; AAY84725.
XX
PT   New isolated polyketide synthase nucleic acid and polyketide compounds,
PT   useful for treating e.g. transplant rejection, uveitis, multiple
PT   sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT   peripheral neuropathy  -
XX
PS   Example 1; Page 76-79; 126pp; English.
XX
CC   The present sequence represents an AvrII-XhoII fragment that encodes
CC   module 8 of te FK-520 polyketide synthase (PKS) gene cluster with the
CC   endogenous acyltransferase (AT) domain replaced by the AT domain of
CC   module 12 of the rapamycin PKS (which is specific for malonyl units).
CC   FK-506 is a potent immunosuppressants, and acts through intial formation
CC   of an intermediate complex with protein immunophilins known as FK-506
CC   binding proteins. The nucleic acids are used for producing polyketide
CC   compounds. The polyketide compounds can be used as immunosuppressants to
CC   prevent or treat transplant rejection, graft-versus-host disease or
CC   uveitis. They can also be used for treating e.g. alopecia universalis,
CC   autoimmune chronic active hepatitis, inflammatory bowel disease,
CC   multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
CC   also have neurotrophic activity and can be used to promote neurite
CC   outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC   and in intact animals, they promote regrowth of damaged facial and
CC   sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC   the brain. They can also be used for treating e.g. Parkinson's disease,
CC   Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC   peripheral neuropathies. They can also be used in agricultural and
CC   veterinary applications.
XX
SQ   Sequence 4478 BP; 632 A; 1757 C; 1465 G; 624 T; 0 other;

  Query Match             45.7%;  Score 244;  DB 21;  Length 4478;
  Best Local Similarity   67.7%;  Pred. No. 3.1e-31;
  Matches  359;  Conservative    0;  Mismatches  165;  Indels    6;  Gaps    1;

Qy          5 CCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCA 64
              || || || |||| |  | || | || | || | ||| |   |||  | ||| ||
Db       3850 CCCTCCACACCACCACCCCACCCACCACCACCCCCCTCAACCCCGAACACGCCATCATCA 3909

Qy         65 CCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACG 124
               |  |    | ||      |||| |  || |  |      |  ||   ||  |||||||
Db       3910 TCACCGGCGGCTCCGGCACCCTCGCCGGCATCCTCGCCCGCCACCTGAACCACCCCCACA 3969

Qy        125 CACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCG 184
               |  | |||| ||  |   |  |  |   | | || || |  ||||  |  | ||||||
Db       3970 CCTACCTCCTCTCCCGCACCCCACCCCCCGACGCCACCCCCGGCACCCACCTCCCCTGCG 4029

Qy        185 ACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCC 244
              ||  | || ||||  ||||||||          |||||  |  |||||||       || ||||
Db       4030 ACGTCGGCGACCCCCACCAACTCGCCACCACCCTCACCCACATCCCC------CAACCCC 4083

Qy        245 TCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCC 304
              ||||| || || |||||||||| | || |||||||||| || ||||  |  |||||||||
Db       4084 TCACCGCCATCTTCCACACCGCCGCCACCCTCGACGACGGCATCCTCCACGCCCTCACCC 4143

Qy        305 CCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAAC 364
              ||  ||  |||| || |||||||| | |||||||| ||  ||||   | ||| ||||| |
Db       4144 CCGACCGCCTCACCACCGTCCTCCACCCCAAAGCCAACGCCGCCTGGCACCTGCACCACC 4203

Qy        365 TCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCG 424
              ||||||| ||   |||||||||  |||||||||||||||  ||||||||||  | |||
Db       4204 TCACCCAAAACCAACCCCTCACCCACTTCGTCCTCTACTCCAGCGCCGCCGCCGTCCTCG 4263

Qy        425 GCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACC 484
              ||   ||||| ||||  |||||||||| |||||||||| ||||||||||||||||  ||
Db       4264 GCAGCCCCGGACAAGGAAACTACGCCGCCGCCAACGCCTTCCTCGACGCCCTCGCCACCC 4323

Qy        485 ACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              ||||||||||||| |  ||  |||||||||  ||||||||||||||  |||
Db       4324 ACCGCCACACCCTCGGCCAACCCGCCACCTCCATCGCCTGGGGCATGTGG 4373


RESULT 14
AAA14664
ID   AAA14664 standard; DNA; 4547 BP.
XX
AC   AAA14664;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Nucleotide sequence of a fragment of the FK-520 PKS gene cluster.
XX
KW   FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW   immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW   transplant rejection; graft-versus-host disease; alopecia universalis;
KW   autoimmune chronic active hepatitis; inflammatory bowel disease;
KW   multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW   neurite outgrowth; nerve regrowth; Parkinson's disease;
KW   Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW   peripheral neuropathy; ss.
XX
OS   Synthetic.
OS   Streptomyces hygroscopicus.
XX
FH   Key             Location/Qualifiers
FT   CDS             9..4535
FT                   /*tag=  a
FT                   /note= "no termination codon given"
XX
PN   WO200020601-A2.
XX
PD   13-APR-2000.
XX
PF   01-OCT-1999;   99WO-US22886.
XX
PR   02-OCT-1998;   98US-0102748.
PR   11-MAR-1999;   99US-0123810.
PR   17-JUN-1999;   99US-0139650.
XX
```

```
Qy        184 GACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCC 243
              |  |  |  | | ||||    ||||      ||     |  || || |  || ||| |||
Db       4371 CCCTGCGACCTCACCGACCCCACCCAAATCACCCAAGCCCTCACCCACATACCACAACCC 4430

Qy        244 CTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACC 303
              ||||||  | ||  |||||||||| | || ||||||||||||||||||||||||||||||
Db       4431 CTCACCGGCATCTTCCACACCGCCGCCACCCTCGACGACGCCACCCTCACCAACCTCACC 4490

Qy        304 CCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAA 363
              |||   || |||| || |  |||||   |||||||| ||  ||||   | |||||||||
Db       4491 CCCCAACACCTCACCACCACCCTCCAACCCAAAGCCGACGCCGCCTGGCACCTCCACCAC 4550

Qy        364 CTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTC 423
              | ||||||| ||   |||||||||  ||||||||||||||||  |||||||||||| ||
Db       4551 CACACCCAAAACCAACCCCTCACCCACTTCGTCCTCTACTCCAGCGCCGCCGCCACCCTC 4610

Qy        424 GGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCAC 483
              |||   |||||||||||||||||||||| |||||||||  |||||||||||||||  |
Db       4611 GGCAGCCCCGGCCAAGCCAACTACGCCGCCGCCAACGCCTTCCTCGACGCCCTCGCCACC 4670

Qy        484 CACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              |||||||||||||   |  |||||||||  ||||||||||||   |||
Db       4671 CACCGCCACACCCAAGGACAACCCGCCACCACCATCGCCTGGGGCATGTGG 4721
```

RESULT 12
AAA14663
ID AAA14663 standard; DNA; 4466 BP.
XX
AC AAA14663;
XX
DT 08-AUG-2000 (first entry)
XX
DE Nucleotide sequence of a fragment of the FK-520 PKS gene cluster.
XX
KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW transplant rejection; graft-versus-host disease; alopecia universalis;
KW autoimmune chronic active hepatitis; inflammatory bowel disease;
KW multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW neurite outgrowth; nerve regrowth; Parkinson's disease;
KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW peripheral neuropathy; ss.
XX
OS Synthetic.
OS Streptomyces hygroscopicus.
XX
FH Key Location/Qualifiers
FT CDS 9..4454
FT /*tag= a
FT /note= "no termination codon given"
XX
PN WO200020601-A2.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-US22886.
XX
PR 02-OCT-1998; 98US-0102748.
PR 11-MAR-1999; 99US-0123810.
PR 17-JUN-1999; 99US-0139650.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Reeves C, Chu D, Khosla C, Santi D, Wu K;
XX
DR WPI; 2000-317716/27.
DR P-PSDB; AAY84727.
XX
PT New isolated polyketide synthase nucleic acid and polyketide compounds,
PT useful for treating e.g. transplant rejection, uveitis, multiple
PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT peripheral neuropathy -
XX
PS Example 1; Page 82-85; 126pp; English.
XX
CC The present sequence represents an NheII-XhoII fragment that encodes
CC module 8 of te FK-520 polyketide synthase (PKS) gene cluster with the
CC endogenous acyltransferase (AT) domain replaced by the AT domain of
CC module 12 of the rapamycin PKS (which is specific for malonyl CoA).
CC FK-506 is a potent immunosuppressants, and acts through intial formation
CC of an intermediate complex with protein immunophilins known as FK-506
CC binding proteins. The nucleic acids are used for producing polyketide
CC compounds. The polyketide compounds can be used as immunosuppressants to
CC prevent or treat transplant rejection, graft-versus-host disease or
CC uveitis. They can also be used for treating e.g. alopecia universalis,
CC autoimmune chronic active hepatitis, inflammatory bowel disease,
CC multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
CC also have neurotrophic activity and can be used to promote neurite
CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC and in intact animals, they promote regrowth of damaged facial and
CC sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC the brain. They can also be used for treating e.g. Parkinson's disease,
CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC peripheral neuropathies. They can also be used in agricultural and
CC veterinary applications.
XX
SQ Sequence 4466 BP; 634 A; 1796 C; 1433 G; 603 T; 0 other;

```
  Query Match             45.7%;  Score 244;  DB 21;  Length 4466;
  Best Local Similarity   67.7%;  Pred. No. 3.1e-31;
  Matches  359;  Conservative    0;  Mismatches  165;  Indels    6;  Gaps    1;

Qy          5 CCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCA 64
              || || || ||||  |  | ||  | ||  | ||  | ||| |   |||  | ||| ||
Db       3838 CCCTCCACACCACCACCCCACCCACCACCACCCCCCTCAACCCCGAACACGCCATCATCA 3897

Qy         65 CCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACG 124
              |  |      | ||      ||||  |  || |  |      |  ||   ||  |||||||
Db       3898 TCACCGGCGGCTCCGGCACCCTCGCCGGCATCCTCGCCCGCCACCTGAACCACCCCCACA 3957

Qy        125 CACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCG 184
              |  | |||| ||  |    |  |    | | || || |  |||| |  | ||||||
Db       3958 CCTACCTCCTCTCCCGCACCCCACCCCCGACGCCACCCCCGGCACCCACCTCCCCTGCG 4017

Qy        185 ACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCC 244
              ||  | || ||||  ||||||||          ||||| |  |||||||      || ||||
Db       4018 ACGTCGGCGACCCCCACCAACTCGCCACCACCCTCACCCACATCCCC------CAACCCC 4071

Qy        245 TCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCC 304
              ||||| || || |||||||||| | || |||||||||| || ||||  |  ||||||||
Db       4072 TCACCGCCATCTTCCACACCGCCGCCACCCTCGACGACGGCATCCTCCACGCCCTCACCC 4131

Qy        305 CCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAAC 364
              ||  || |||| || |||||||| | |||||||| ||  ||||   | ||| ||||| |
Db       4132 CCGACCGCCTCACCACCGTCCTCCACCCCAAAGCCAACGCCGCCTGGCACCTGCACCACG 4191

Qy        365 TCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCG 424
              |||||||| ||   |||||||||  |||||||||||||||  ||||||||||  | |||
Db       4192 TCACCCAAAACCAACCCCTCACCCACTTCGTCCTCTACTCCAGCGCCGCCGCCGTCCTCG 4251

Qy        425 GCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACC 484
              ||   ||||| ||||  |||||||||| |||||||||| ||||||||||||||  ||
Db       4252 GCAGCCCCGGACAAGGAAACTACGCCGCCGCCAACGCCTTCCTCGACGCCCTCGCCACCC 4311

Qy        485 ACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              |||||||||||| |  ||  ||||||||  |||||||||||||  |||
Db       4312 ACCGCCACACCCTCGGCCAACCCGCCACCTCCATCGCCTGGGGCATGTGG 4361
```

RESULT 13
AAA14661
ID AAA14661 standard; DNA; 4478 BP.
XX
AC AAA14661;
XX

```
CC   compounds can be used as immunosuppressants to prevent or treat
CC   transplant rejection, graft-versus-host disease or uveitis. They can
CC   also be used for treating e.g. alopecia universalis, autoimmune
CC   chronic active hepatitis, inflammatory bowel disease, multiple
CC   sclerosis, primary biliary cirrhosis, or scleroderma. They
CC   also have neurotrophic activity and can be used to promote neurite
CC   outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC   and in intact animals, they promote regrowth of damaged facial and
CC   sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC   the brain. They can also be used for treating e.g. Parkinson's disease,
CC   Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC   peripheral neuropathies. They can also be used in agricultural and
CC   veterinary applications.
XX
SQ   Sequence 4767 BP; 731 A; 1945 C; 1468 G; 623 T; 0 other;

  Query Match             48.2%;  Score 257.4;  DB 21;  Length 4767;
  Best Local Similarity   67.8%;  Pred. No. 2.1e-33;
  Matches  360;  Conservative    0;  Mismatches  171;  Indels    0;  Gaps    0;

Qy          4 ACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACC 63
              |||  || || ||||  |  ||||  | ||  | || | | ||||| || | ||| | ||
Db       4140 ACCCACCACAACACCACCACAACCACCCCCAACACCCCACCCCTCAACCCCAACCACGCC 4199

Qy         64 ACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCAC 123
              | || |    ||  |     |  |  |||| || ||     || || ||||  |  ||||
Db       4200 ATCCTCATCACCGGCGGCTCCGGCACCCTCGCCGGCATCCTCGCCCGCCACCTCAACCAC 4259

Qy        124 GCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGC 183
               | ||   |  |  |  |  | |  || |  |  |  |  |    ||  ||||  |  |
Db       4260 CCCCACACCTACCTCCTCTCCCGCACACCACCACCCCCCACCACACCCGGCACCCACATC 4319

Qy        184 GACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCC 243
               |  |  |  | | ||||    ||||      ||    |  || || |  || ||| |||
Db       4320 CCCTGCGACCTCACCGACCCCACCCAAATCACCCAAGCCCTCACCCACATACCACAACCC 4379

Qy        244 CTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACC 303
              ||||||  | || |||||||||| | || ||||||||||||||||||||||||||||||
Db       4380 CTCACCGGCATCTTCCACACCGCCGCCACCCTCGACGACGCCACCCTCACCAACCTCACC 4439

Qy        304 CCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAA 363
              |||   || |||| || |  |||||   |||||||| ||  ||||   | |||||||||
Db       4440 CCCCAACACCTCACCACCACCCTCCAACCCAAAGCCGACGCCGCCTGGCACCTCCACCAC 4499

Qy        364 CTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTC 423
              | ||||||| ||   |||||||||  |||||||||||||||  |||||||||||| ||
Db       4500 CACACCCAAAACCAACCCCTCACCCACTTCGTCCTCTACTCCAGCGCCGCCGCCACCCTC 4559

Qy        424 GGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCAC 483
              |||   ||||||||||||||||||||| |||||||||| ||||||||||||||||   |
Db       4560 GGCAGCCCCGGCCAAGCCAACTACGCCGCCGCCAACGCCTTCCTCGACGCCCTCGCCACC 4619

Qy        484 CACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              |||||||||||||    |  |||||||||| |||||||||||||  |||
Db       4620 CACCGCCACACCCAAGGACAACCCGCCACCACCATCGCCTGGGGCATGTGG 4670


RESULT 11
AAA14669
ID   AAA14669 standard; DNA; 4818 BP.
XX
AC   AAA14669;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Nucleotide sequence of modified FK-520 PKS gene cluster module 8.
XX
KW   FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW   immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW   transplant rejection; graft-versus-host disease; alopecia universalis;
KW   autoimmune chronic active hepatitis; inflammatory bowel disease;
KW   multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW   neurite outgrowth; nerve regrowth; Parkinson's disease;
KW   Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW   peripheral neuropathy; ss.
XX
OS   Synthetic.
OS   Streptomyces hygroscopicus.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..4817
FT                   /*tag=  a
FT                   /note= "no termination codon given"
XX
PN   WO200020601-A2.
XX
PD   13-APR-2000.
XX
PF   01-OCT-1999;   99WO-US22886.
XX
PR   02-OCT-1998;   98US-0102748.
PR   11-MAR-1999;   99US-0123810.
PR   17-JUN-1999;   99US-0139650.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Reeves C,  Chu D,  Khosla C,  Santi D,  Wu K;
XX
DR   WPI; 2000-317716/27.
DR   P-PSDB; AAY84733.
XX
PT   New isolated polyketide synthase nucleic acid and polyketide compounds,
PT   useful for treating e.g. transplant rejection, uveitis, multiple
PT   sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT   peripheral neuropathy  -
XX
PS   Example 2; Page 102-105; 126pp; English.
XX
CC   The present sequence represents module 8 of the FK-520 polyketide
CC   synthase (PKS) gene cluster, containing the acyltransferase (AT)
CC   domain of module 13 of rapamycin. FK-506 is a potent immunosuppressant,
CC   and acts through intial formation of an intermediate complex with
CC   protein immunophilins known as FK-506 binding proteins. The nucleic
CC   acids are used for producing polyketide compounds. The polyketide
CC   compounds can be used as immunosuppressants to prevent or treat
CC   transplant rejection, graft-versus-host disease or uveitis. They can
CC   also be used for treating e.g. alopecia universalis, autoimmune
CC   chronic active hepatitis, inflammatory bowel disease, multiple
CC   sclerosis, primary biliary cirrhosis, or scleroderma. They
CC   also have neurotrophic activity and can be used to promote neurite
CC   outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC   and in intact animals, they promote regrowth of damaged facial and
CC   sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC   the brain. They can also be used for treating e.g. Parkinson's disease,
CC   Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC   peripheral neuropathies. They can also be used in agricultural and
CC   veterinary applications.
XX
SQ   Sequence 4818 BP; 742 A; 1982 C; 1476 G; 618 T; 0 other;

  Query Match             48.2%;  Score 257.4;  DB 21;  Length 4818;
  Best Local Similarity   67.8%;  Pred. No. 2.1e-33;
  Matches  360;  Conservative    0;  Mismatches  171;  Indels    0;  Gaps    0;

Qy          4 ACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACC 63
              |||  || || ||||  |  ||||  | ||  | || | | ||||| || | ||| | ||
Db       4191 ACCCACCACAACACCACCACAACCACCCCCAACACCCCACCCCTCAACCCCAACCACGCC 4250

Qy         64 ACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCAC 123
              | || |    ||  |     |  |  |||| || ||     || || ||||  |  ||||
Db       4251 ATCCTCATCACCGGCGGCTCCGGCACCCTCGCCGGCATCCTCGCCCGCCACCTCAACCAC 4310

Qy        124 GCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGC 183
               | ||   |  |  |  |  | |  || |  |  |  |  |    ||  ||||  |  |
Db       4311 CCCCACACCTACCTCCTCTCCCGCACACCACCACCCCCCACCACACCCGGCACCCACATC 4370
```

```
FT                   /note= "no termination codon given"
XX
PN   WO200020601-A2.
XX
PD   13-APR-2000.
XX
PF   01-OCT-1999;   99WO-US22886.
XX
PR   02-OCT-1998;   98US-0102748.
PR   11-MAR-1999;   99US-0123810.
PR   17-JUN-1999;   99US-0139650.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Reeves C,  Chu D,  Khosla C,  Santi D,  Wu K;
XX
DR   WPI; 2000-317716/27.
DR   P-PSDB; AAY84732.
XX
PT   New isolated polyketide synthase nucleic acid and polyketide compounds,
PT   useful for treating e.g. transplant rejection, uveitis, multiple
PT   sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT   peripheral neuropathy  -
XX
PS   Example 2; Page 99-102; 126pp; English.
XX
CC   The present sequence represents module 8 of the FK-520 polyketide
CC   synthase (PKS) gene cluster, containing the acyltransferase (AT)
CC   domain of module 12 of rapamycin. FK-506 is a potent immunosuppressant,
CC   and acts through intial formation of an intermediate complex with
CC   protein immunophilins known as FK-506 binding proteins. The nucleic
CC   acids are used for producing polyketide compounds. The polyketide
CC   compounds can be used as immunosuppressants to prevent or treat
CC   transplant rejection, graft-versus-host disease or uveitis. They can
CC   also be used for treating e.g. alopecia universalis, autoimmune
CC   chronic active hepatitis, inflammatory bowel disease, multiple
CC   sclerosis, primary biliary cirrhosis, or scleroderma. They
CC   also have neurotrophic activity and can be used to promote neurite
CC   outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC   and in intact animals, they promote regrowth of damaged facial and
CC   sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC   the brain. They can also be used for treating e.g. Parkinson's disease,
CC   Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC   peripheral neuropathies. They can also be used in agricultural and
CC   veterinary applications.
XX
SQ   Sequence 4737 BP; 718 A; 1927 C; 1472 G; 620 T; 0 other;

  Query Match             48.2%;  Score 257.4;  DB 21;  Length 4737;
  Best Local Similarity   67.8%;  Pred. No. 2.1e-33;
  Matches  360;  Conservative    0;  Mismatches  171;  Indels    0;  Gaps    0;

Qy          4 ACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACC 63
              |||  || || ||||  |  ||||  | ||  | || | | ||||| || | ||| | ||
Db       4110 ACCCACCACAACACCACCACAACCACCCCCAACACCCCACCCCTCAACCCCAACCACGCC 4169

Qy         64 ACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCAC 123
              | || |    ||  |     |  |  |||| || ||     || || ||||  |  ||||
Db       4170 ATCCTCATCACCGGCGGCTCCGGCACCCTCGCCGGCATCCTCGCCCGCCACCTCAACCAC 4229

Qy        124 GCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGC 183
               | ||   |  |  |  |   | |  || | |  |  |  |     ||  ||||  |  |
Db       4230 CCCCACACCTACCTCCTCTCCCGCACACCACCACCCCCCCACCACACCCGGCACCCACATC 4289

Qy        184 GACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCC 243
              |  |  |  | | ||||    ||||      ||     |  || || |  || ||| |||
Db       4290 CCCTGCGACCTCACCGACCCCACCCAAATCACCCAAGCCCTCACCCACATACCACAACCC 4349

Qy        244 CTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACC 303
              ||||||  | || |||||||||| | || ||||||||||||||||||||||||||||||
Db       4350 CTCACCGGCATCTTCCACACCGCCGCCACCCTCGACGACGCCACCCTCACCAACCTCACC 4409

Qy        304 CCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAA 363
              |||   || |||| || |  |||||   |||||||| ||  ||||   | |||||||||
Db       4410 CCCCAACACCTCACCACCACCCTCCAACCCAAAGCCGACGCCGCCTGGCACCTCCACCAC 4469

Qy        364 CTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTC 423
              | ||||||| ||   |||||||||  ||||||||||||||||  |||||||||||| ||
Db       4470 CACACCCAAAACCAACCCCTCACCCACTTCGTCCTCTACTCCAGCGCCGCCGCCACCCTC 4529

Qy        424 GGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCAC 483
              |||   |||||||||||||||||||||| |||||||||| ||||||||||||||||  |
Db       4530 GGCAGCCCCGGCCAAGCCAACTACGCCGCCGCCAACGCCTTCCTCGACGCCCTCGCCACC 4589

Qy        484 CACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              |||||||||||||    | |||||||||| ||||||||||||  |||
Db       4590 CACCGCCACACCCAAGGACAACCCGCCACCACCATCGCCTGGGGCATGTGG 4640


RESULT 10
AAA14667
ID   AAA14667 standard; DNA; 4767 BP.
XX
AC   AAA14667;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Nucleotide sequence of modified FK-520 PKS gene cluster module 8.
XX
KW   FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW   immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW   transplant rejection; graft-versus-host disease; alopecia universalis;
KW   autoimmune chronic active hepatitis; inflammatory bowel disease;
KW   multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW   neurite outgrowth; nerve regrowth; Parkinson's disease;
KW   Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW   peripheral neuropathy; ss.
XX
OS   Synthetic.
OS   Streptomyces hygroscopicus.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..4766
FT                   /*tag=  a
FT                   /note= "no termination codon given"
XX
PN   WO200020601-A2.
XX
PD   13-APR-2000.
XX
PF   01-OCT-1999;   99WO-US22886.
XX
PR   02-OCT-1998;   98US-0102748.
PR   11-MAR-1999;   99US-0123810.
PR   17-JUN-1999;   99US-0139650.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Reeves C,  Chu D,  Khosla C,  Santi D,  Wu K;
XX
DR   WPI; 2000-317716/27.
DR   P-PSDB; AAY84731.
XX
PT   New isolated polyketide synthase nucleic acid and polyketide compounds,
PT   useful for treating e.g. transplant rejection, uveitis, multiple
PT   sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT   peripheral neuropathy  -
XX
PS   Example 2; Page 96-99; 126pp; English.
XX
CC   The present sequence represents module 8 of the FK-520 polyketide
CC   synthase (PKS) gene cluster, containing the acyltransferase (AT)
CC   domain of module 13 of rapamycin. FK-506 is a potent immunosuppressant,
CC   and acts through intial formation of an intermediate complex with
CC   protein immunophilins known as FK-506 binding proteins. The nucleic
CC   acids are used for producing polyketide compounds. The polyketide
```

```
Qy        424 GGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCAC 483
              |||   |||||||||||||||||||||| |||||||||| ||||||||||||||||  |
Db       4467 GGCAGCCCCGGCCAAGCCAACTACGCCGCCGCCAACGCCTTCCTCGACGCCCTCGCCACC 4526

Qy        484 CACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              |||||||||||||    |  |||||||||| |||||||||||||  |||
Db       4527 CACCGCCACACCCAAGGACAACCCGCCACCACCATCGCCTGGGGCATGTGG 4577
```

RESULT 8
AAA14665
ID AAA14665 standard; DNA; 4725 BP.
XX
AC AAA14665;
XX
DT 08-AUG-2000 (first entry)
XX
DE Nucleotide sequence of FK-520 PKS gene cluster module 8.
XX
KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW transplant rejection; graft-versus-host disease; alopecia universalis;
KW autoimmune chronic active hepatitis; inflammatory bowel disease;
KW multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW neurite outgrowth; nerve regrowth; Parkinson's disease;
KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW peripheral neuropathy; ss.
XX
OS Streptomyces hygroscopicus.
XX
FH Key Location/Qualifiers
FT CDS 3..4724
FT /*tag= a
FT /note= "no termination codon given"
XX
PN WO200020601-A2.
XX
PD 13-APR-2000.
XX
PF 01-OCT-1999; 99WO-US22886.
XX
PR 02-OCT-1998; 98US-0102748.
PR 11-MAR-1999; 99US-0123810.
PR 17-JUN-1999; 99US-0139650.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Reeves C, Chu D, Khosla C, Santi D, Wu K;
XX
DR WPI; 2000-317716/27.
DR P-PSDB; AAY84729.
XX
PT New isolated polyketide synthase nucleic acid and polyketide compounds,
PT useful for treating e.g. transplant rejection, uveitis, multiple
PT sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT peripheral neuropathy -
XX
PS Example 2; Page 90-93; 126pp; English.
XX
CC The present sequence encodes module 8 of the FK-520 polyketide
CC synthase (PKS) gene cluster of strain MA6548. FK-506 is a potent
CC immunosuppressants, and acts through intial formation of an
CC intermediate complex with protein immunophilins known as FK-506
CC binding proteins. The nucleic acids are used for producing polyketide
CC compounds. The polyketide compounds can be used as immunosuppressants to
CC prevent or treat transplant rejection, graft-versus-host disease or
CC uveitis. They can also be used for treating e.g. alopecia universalis,
CC autoimmune chronic active hepatitis, inflammatory bowel disease,
CC multiple sclerosis, primary biliary cirrhosis, or scleroderma. They
CC also have neurotrophic activity and can be used to promote neurite
CC outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC and in intact animals, they promote regrowth of damaged facial and
CC sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC the brain. They can also be used for treating e.g. Parkinson's disease,
CC Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC peripheral neuropathies. They can also be used in agricultural and
CC veterinary applications.
XX
SQ Sequence 4725 BP; 728 A; 2034 C; 1394 G; 569 T; 0 other;

```
  Query Match             48.2%;  Score 257.4;  DB 21;  Length 4725;
  Best Local Similarity   67.8%;  Pred. No. 2.1e-33;
  Matches  360;  Conservative    0;  Mismatches  171;  Indels    0;  Gaps    0;

Qy          4 ACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACC 63
              |||  || || ||||  |  ||||  | ||  | || | | ||||| || | ||| | ||
Db       4098 ACCCACCACAACACCACCACAACCACCCCCAACACCCCACCCCTCAACCCCAACCACGCC 4157

Qy         64 ACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCAC 123
              | || |    ||  |     |  |  |||| || ||     || || ||||  |  ||||
Db       4158 ATCCTCATCACCGGCGGCTCCGGCACCCTCGCCGGCATCCTCGCCCGCCACCTCAACCAC 4217

Qy        124 GCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGC 183
              | |||   |  |  |  |   | |  || | |  |  |  |  |     ||  |||| |  |
Db       4218 CCCCACACCTACCTCCTCTCCCGCACACCACCACCCCCCACCACACCCGGCACCCACATC 4277

Qy        184 GACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCC 243
              |  |  |  | | ||||    ||||    ||    |  || || |  || ||| |||
Db       4278 CCCTGCGACCTCACCGACCCCACCCAAATCACCCAAGCCCTCACCCACATACCACAACCC 4337

Qy        244 CTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACC 303
              |||||| | || |||||||||| | || ||||||||||||||||||||||||||||||||
Db       4338 CTCACCGGCATCTTCCACACCGCCGCCACCCTCGACGACGCCACCCTCACCAACCTCACC 4397

Qy        304 CCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAA 363
              |||   || |||| || |  |||||   |||||||| ||  ||||   | |||||||||
Db       4398 CCCCAACACCTCACCACCACCCTCCAACCCAAAGCCGACGCCGCCTGGCACCTCCACCAC 4457

Qy        364 CTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTC 423
              | ||||||| ||   |||||||||  |||||||||||||||  ||||||||||||| ||
Db       4458 CACACCCAAAACCAACCCCTCACCCACTTCGTCCTCTACTCCAGCGCCGCCGCCACCCTC 4517

Qy        424 GGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCAC 483
              |||   |||||||||||||||||||||| |||||||||| ||||||||||||||||  |
Db       4518 GGCAGCCCCGGCCAAGCCAACTACGCCGCCGCCAACGCCTTCCTCGACGCCCTCGCCACC 4577

Qy        484 CACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              |||||||||||||    |  |||||||||| |||||||||||||  |||
Db       4578 CACCGCCACACCCAAGGACAACCCGCCACCACCATCGCCTGGGGCATGTGG 4628
```

RESULT 9
AAA14668
ID AAA14668 standard; DNA; 4737 BP.
XX
AC AAA14668;
XX
DT 08-AUG-2000 (first entry)
XX
DE Nucleotide sequence of modified FK-520 PKS gene cluster module 8.
XX
KW FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW transplant rejection; graft-versus-host disease; alopecia universalis;
KW autoimmune chronic active hepatitis; inflammatory bowel disease;
KW multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW neurite outgrowth; nerve regrowth; Parkinson's disease;
KW Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW peripheral neuropathy; ss.
XX
OS Synthetic.
OS Streptomyces hygroscopicus.
XX
FH Key Location/Qualifiers
FT CDS 3..4736
FT /*tag= a

```
  Query Match             48.8%;  Score 260.4;  DB 22;  Length 65140;
  Best Local Similarity   68.0%;  Pred. No. 6e-34;
  Matches  363;  Conservative    0;  Mismatches  171;  Indels    0;  Gaps    0;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  ||| |||||||||||| ||||||| ||| | | |   |||| ||| ||||||
Db      20690 GGCACCGTCCTGATCACCGGCGGCACCGGCGGCCTGGGCGCGGTCCTCGCCCGCCACCTG 20749

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
                | ||  |||    |  ||  ||||| || ||| |||||||   ||||||   | ||  |
Db      20750 GTCGCCACCCACGGCGTCCGCCACCTGCTGCTCGCCAGCCGCCCGCGGCCCGGCCGCCGAC 20809

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
               |||  |  |||| ||  || |||||          |||  |  | ||  |||| || ||
Db      20810 GGCGCCGACGACCTGACGGCCGAACTCACCGGGCTCGGCGCCACCGTCCACATCGCCGCC 20869

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||  |  | |||| | |   |||    | || |||  |||| |||| |    |||
Db      20870 TGCGACGTCGCCGACCCGGCCGCCCTCGCCGACCTGCTCGGCACCGTCCCGGCCGGGCAC 20929

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              || ||||||  |||| |||||||||| ||| || |||||||| |  ||||  |  ||||
Db      20930 CCGCTCACCGTCGTCGTCCACACCGCCGGCGTCGTCGACGACGGCGTCCTCGGCTCCCTC 20989

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              |||||    |  ||  ||| |||||| ||  |||| ||| ||  ||||   | ||| |||
Db      20990 ACCCCGCAGCGCCTGGACACCGTCCTGCGGCCCAAGGCCGACGCCGCCTGGCACCTGCAC 21049

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
               |   ||||  |||  |  |||   |||||||||||||| ||| |||| ||||||||||
Db      21050 GAGGCGACCCGCCACCTCGACCTGGACGCCTTCGTCCTCTTCTCGTCCGTCGCCGCCACC 21109

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              |||||   ||||| || |||||||||||||| | |||||||| ||| ||||||||||||
Db      21110 CTCGGCAGCCCCGGACAGGCCAACTACGCCGCCGGCAACGCCTTCCTGGACGCCCTCGCC 21169

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              ||  |||  | ||  |  ||| |||||||||  | ||||||||||| | |||
Db      21170 GCCCGGCGCGCCGCCACCGGCCTGCCCGCCACCTCCCTCGCCTGGGGCCCGTGG 21223


RESULT 7
AAA14666
ID   AAA14666 standard; DNA; 4674 BP.
XX
AC   AAA14666;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Nucleotide sequence of modified FK-520 PKS gene cluster module 8.
XX
KW   FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW   immunophilin; FK-506 binding protein; polyketide compound; uveitis;
KW   transplant rejection; graft-versus-host disease; alopecia universalis;
KW   autoimmune chronic active hepatitis; inflammatory bowel disease;
KW   multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW   neurite outgrowth; nerve regrowth; Parkinson's disease;
KW   Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW   peripheral neuropathy; ss.
XX
OS   Synthetic.
OS   Streptomyces hygroscopicus.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..4673
FT                   /*tag=  a
FT                   /note= "no termination codon given"
XX
PN   WO200020601-A2.
XX
PD   13-APR-2000.
XX
PF   01-OCT-1999;   99WO-US22886.
XX
PR   02-OCT-1998;   98US-0102748.
PR   11-MAR-1999;   99US-0123810.
PR   17-JUN-1999;   99US-0139650.
XX
PA   (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI   Reeves C,  Chu D,  Khosla C,  Santi D,  Wu K;
XX
DR   WPI; 2000-317716/27.
DR   P-PSDB; AAY84730.
XX
PT   New isolated polyketide synthase nucleic acid and polyketide compounds,
PT   useful for treating e.g. transplant rejection, uveitis, multiple
PT   sclerosis, Alzheimer's disease, Parkinson's disease, stroke, or
PT   peripheral neuropathy  -
XX
PS   Example 2; Page 93-96; 126pp; English.
XX
CC   The present sequence represents module 8 of the FK-520 polyketide
CC   synthase (PKS) gene cluster, containing the acyltransferase (AT)
CC   domain of module 12 of rapamycin. FK-506 is a potent immunosuppressant,
CC   and acts through intial formation of an intermediate complex with
CC   protein immunophilins known as FK-506 binding proteins. The nucleic
CC   acids are used for producing polyketide compounds. The polyketide
CC   compounds can be used as immunosuppressants to prevent or treat
CC   transplant rejection, graft-versus-host disease or uveitis. They can
CC   also be used for treating e.g. alopecia universalis, autoimmune
CC   chronic active hepatitis, inflammatory bowel disease, multiple
CC   sclerosis, primary biliary cirrhosis, or scleroderma. They
CC   also have neurotrophic activity and can be used to promote neurite
CC   outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures,
CC   and in intact animals, they promote regrowth of damaged facial and
CC   sciatic nerves, and repair lesioned serotonin and dopamine neurons in
CC   the brain. They can also be used for treating e.g. Parkinson's disease,
CC   Alzheimer's disease, stroke, traumatic spinal cord and brain injury, or
CC   peripheral neuropathies. They can also be used in agricultural and
CC   veterinary applications.
XX
SQ   Sequence 4674 BP; 704 A; 1873 C; 1464 G; 633 T; 0 other;

  Query Match             48.2%;  Score 257.4;  DB 21;  Length 4674;
  Best Local Similarity   67.8%;  Pred. No. 2.1e-33;
  Matches  360;  Conservative    0;  Mismatches  171;  Indels    0;  Gaps    0;

Qy          4 ACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACC 63
              |||  || || |||| | |||| | ||  | || | | ||||| || | ||| | ||
Db       4047 ACCCACCACAACACCACCACAACCACCCCCAACACCCCACCCCTCAACCCCAACCACGCC 4106

Qy         64 ACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCAC 123
              | ||| |    || |     | |  |||| || ||     || || |||| |  ||||
Db       4107 ATCCTCATCACCGGCGGCTCCGGCACCCTCGCCGGCATCCTCGCCCGCCACCTCAACCAC 4166

Qy        124 GCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGC 183
               | ||  |  |  |  |  | | || | |  |  |  |    ||  |||| |  |
Db       4167 CCCCACACCTACCTCCTCTCCCGCACACCACCACCCCCCACCACACCCGGCACCCACATC 4226

Qy        184 GACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCC 243
               |  |  |  | | ||||   ||||     ||    |  || || |  || ||| |||
Db       4227 CCCTGCGACCTCACCGACCCCACCCAAATCACCCAAGCCCTCACCCACATACCACAACCC 4286

Qy        244 CTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACC 303
              ||||||  | || |||||||||| | || ||||||||||||||||||||||||||||||
Db       4287 CTCACCGGCATCTTCCACACCGCCGCCACCCTCGACGACGCCACCCTCACCAACCTCACC 4346

Qy        304 CCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAA 363
              |||   || |||| || |  |||||   |||||||| ||  ||||   | |||||||||
Db       4347 CCCCAACACCTCACCACCACCCTCCAACCCAAAGCCGACGCCGCCTGGCACCTCCACCAC 4406

Qy        364 CTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTC 423
              | ||||||| ||   |||||||||  |||||||||||||||  ||||||||||||| ||
Db       4407 CACACCCAAAACCAACCCCTCACCCACTTCGTCCTCTACTCCAGCGCCGCCGCCACCCTC 4466
```

```
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match             50.0%;  Score 266.8;  DB 22;  Length 125401;
  Best Local Similarity   68.7%;  Pred. No. 5.4e-35;
  Matches  367;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||   ||| |||||||||| ||||||| ||| |||  ||  ||| ||| ||||||
Db      40009 GGCACCGTGCTCCTCACCGGCGGCACCGGCGGCCTGGCCGGCCTGCTCGCCCGCCACCTG 40068

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
               |||| | ||   | |||  ||||| || ||  || ||||   ||||||| || | |||
Db      40069 GTCACCGAGCACGGCGCCCGCCACCTGCTGCTGGCCGGCCGGCGCGGCCCCGACGCGCCC 40128

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              |||        |||  | || ||||           |||  | |  ||||| || || ||
Db      40129 GGCGCCGCCGCACTCCACGCCGAACTGACCGCCCTGGGCGCCGAGGTCACCGTCGCCGCC 40188

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||  |  | |||     |   |||       || |||  |||| | ||| |  |||||
Db      40189 TGCGACGTCGCCGACCGCACCGCGCTCGCCGCGCTGCTCGCCACCGTGCCCGCCGAACAC 40248

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              ||||||||| | ||| | |||||||| ||| |||| ||||||| ||||||||||  |||
Db      40249 CCCCTCACCGCGGTCGTGCACACCGCCGGCGTCCTGGACGACGGCACCCTCACCGCCCTG 40308

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              | ||||  ||  |||  || |||||| ||  |||| |   ||  ||||    | ||| |||
Db      40309 AACCCCGACCGCCTCGCCACCGTCCTACGGCCCAAGGTGGACGCCGCCTGGCACCTGCAC 40368

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | |||||||  |||  |  |||  |||| ||||| |||||||||||| |||||| |  |
Db      40369 GACCTCACCCGCCACCTCGACCTGGCCGCGTTCGTGCTCTACTCCTCCACCGCCGGCGTC 40428

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              | |||| ||| ||||| |||||||||||| || | |||| ||| |||||||||| ||||||
Db      40429 ATGGGCGGACCGGGCCAGGCCAACTACGCGGCCGGCAACACCTTCCTCGACGCGCTCGCC 40488

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              |||||| ||| ||| |  ||| |||||||||    | |||||||| |||||
Db      40489 GCCCACCGACACGCCCTCGGCCTGCCCGCCACCTCGCTGGCCTGGGGCGCCTGG 40542


RESULT 6
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..1035)
FT                   /*tag=  a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement (1056..2576)
FT                   /*tag=  b
FT                   /product= "NysD1 protein"
FT   CDS             2806..6906
FT                   /*tag=  c
FT                   /product= "NysA protein"
FT   CDS             6952..16530
FT                   /*tag=  d
FT                   /product= "NysB protein"
FT   CDS             16550..49840
FT                   /*tag=  e
FT                   /product= "NysC protein"
FT   CDS             50260..51015
FT                   /*tag=  f
FT                   /product= "NysE protein"
FT   CDS             51405..54305
FT                   /*tag=  g
FT                   /product= "NysR1 protein"
FT   CDS             54329..57190
FT                   /*tag=  h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180..59963
FT                   /*tag=  i
FT                   /product= "NysR3 protein"
FT   CDS             60415..61047
FT                   /*tag=  j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736..62497
FT                   /*tag=  k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (62551..63615)
FT                   /*tag=  l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765..64961
FT                   /*tag=  m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 116-151; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;
```

```
XX
PT   Mutated Type I polyketide synthase containing a polylinker site in an
PT   extension module for replacement of a reductive loop sequence, for
PT   producing polyketides, e.g. B1 avermectin -
XX
PS   Disclosure; Fig 7a-f; 75pp; English.
XX
CC   The present sequence is that of DNA encoding the first 2 modules
CC   of the avermectin polyketide synthase (PKS) of Streptomyces
CC   avermitilis.  The invention relates to nucleic acids encoding a
CC   Type I PKS such as avermectin in which a polylinker with multiple
CC   restriction sites replaces or 1 more PKS genes encoding enzymes
CC   associated with reduction.  Novel PKS are provided in which in
CC   which the reductive loop in a selected module of the Type I PKS is
CC   replaced with the equivalent segment from the same or different
CC   PKS gene cluster or by a mutated or synthetic segment.  Vectors and
CC   host cells, and methods for producing novel polyketides by
CC   culturing host cells are claimed.  The polyketides obtained are
CC   useful as antibiotics and insecticides.  Fermentation products
CC   containing C22-C23 dihydroavermectin, ivermectin and B1
CC   avermectins are claimed.
XX
SQ   Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;

  Query Match             99.7%;  Score 532.4;  DB 21;  Length 12381;
  Best Local Similarity   99.8%;  Pred. No. 9.8e-78;
  Matches  533;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5960 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 6019

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6020 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 6079

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6080 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 6139

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       6140 TGCGACACCAGCAACCCAGACCAACTCCAACATCTCCTCAACACCATCCCCCCACAACAC 6199

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6200 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 6259

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6260 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 6319

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6320 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 6379

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6380 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 6439

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6440 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 6493


RESULT 5
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag= a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag= b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag= c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag= d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag= e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag= f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag= g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag= h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
```

```
Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5443 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 5502

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5503 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 5562

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5563 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 5622

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5623 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 5676
```

RESULT 3
AAH79277
ID AAH79277 standard; DNA; 30690 BP.
XX
AC AAH79277;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis coding sequences SEQ ID NO: 1.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide; ds.
XX
OS Streptomyces avermitilis.
XX

FH	Key	Location/Qualifiers
FT	CDS	1..11919
FT		/*tag= a
FT		/product= "AAG65264"
FT		/partial
FT	CDS	11971..30690
FT		/*tag= b
FT		/product= "AAG65265"

XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX
PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR P-PSDB; AAG65264, AAG65265.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Example 2; Page 58-123; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is a fragment of the S.
CC avermitilis genome.
XX
SQ Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;

Query Match 100.0%; Score 534; DB 22; Length 30690;
Best Local Similarity 100.0%; Pred. No. 5.2e-78;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5143 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 5202

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5203 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 5262

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5263 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 5322

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5323 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 5382

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5383 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 5442

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5443 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 5502

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5503 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 5562

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5563 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 5622

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5623 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 5676
```

RESULT 4
AAZ58381
ID AAZ58381 standard; DNA; 12381 BP.
XX
AC AAZ58381;
XX
DT 23-MAY-2000 (first entry)
XX
DE Streptomyces avermitilis avermectin polyketide synthase modules 1+2.
XX
KW Polyketide synthase; avermectin; insecticide; ss.
XX
OS Streptomyces avermitilis.
XX
PN WO200001827-A2.
XX
PD 13-JAN-2000.
XX
PF 06-JUL-1999; 99WO-GB02158.
XX
PR 06-JUL-1998; 98GB-0014622.
XX
PA (BIOT-) BIOTICA TECHNOLOGY LTD.
PA (PFIZ) PFIZER INC.
XX
PI Kellenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ;
PI McArthur HAI;
XX
DR WPI; 2000-182117/16.

PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR P-PSDB; AAG65268.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 10; Page 149-167; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is modified version of a
CC fragment of the S. avermitilis genome.
XX
SQ Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;

Query Match 100.0%; Score 534; DB 22; Length 11916;
Best Local Similarity 100.0%; Pred. No. 5.4e-78;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5143 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 5202

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5203 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 5262

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5263 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 5322

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5323 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 5382

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5383 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 5442

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5443 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 5502

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5503 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 5562

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5563 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 5622

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5623 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 5676
```

RESULT 2
AAA92301
ID AAA92301 standard; DNA; 30690 BP.
XX
AC AAA92301;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical; ds.
XX
OS Streptomyces avermitilis.
XX
FH Key Location/Qualifiers
FT CDS 1..11919
FT /*tag= a
FT /note= "avermectin aglycon synthase protein"
FT CDS 11971..30690
FT /*tag= b
FT /note= "avermectin aglycon synthase protein"
XX
PN WO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX
PA (KITA) KITASATO INST.
XX
PI Omura S, Ikeda H;
XX
DR WPI; 2000-565458/52.
DR P-PSDB; AAB23749, AAB23750.
XX
PT Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT it for the production of avermectin and its derivatives for drug and
PT agrochemical use -
XX
PS Claim 2; Page 66-134; 314pp; Japanese.
XX
CC The present sequence represents DNA which encodes avermectin aglycon
CC synthase proteins. Also described are: (1) polypeptides encoded by all
CC or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC cells transformed by the vectors; (4) preparation of the polypeptides
CC by culture of the transformants; (5) preparation of avermectin aglycon
CC or its derivatives by culture of transformed avermectin-producing
CC microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC containing sense or antisense sequences from the avermectin aglycon
CC synthase DNA. The enzymes are useful for the production of modified
CC forms of avermectin and of the intermediates in its biosynthesis, for
CC use as drugs, veterinary drugs and agrochemicals.
XX
SQ Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;

Query Match 100.0%; Score 534; DB 21; Length 30690;
Best Local Similarity 100.0%; Pred. No. 5.2e-78;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5143 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 5202

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5203 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 5262

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5263 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 5322

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5323 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 5382

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5383 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 5442
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 106.227 Seconds
(without alignments)
11320.736 Million cell updates/sec

Title: US-09-914-286-1_COPY_5143_5676
Perfect score: 534
Sequence: 1 ggaaccaccctcatcaccgg.....gcatcgcctggggcacctgg 534

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	534	100.0	11916	22	AAH79279	Streptomyces averm
	2	534	100.0	30690	21	AAA92301	S. avermitilis ave
	3	534	100.0	30690	22	AAH79277	Streptomyces averm
	4	532.4	99.7	12381	21	AAZ58381	Streptomyces averm
	5	266.8	50.0	125401	22	AAD17186	Streptomyces nours
	6	260.4	48.8	65140	22	AAD17184	Streptomyces nours
	7	257.4	48.2	4674	21	AAA14666	Nucleotide sequenc
	8	257.4	48.2	4725	21	AAA14665	Nucleotide sequenc
	9	257.4	48.2	4737	21	AAA14668	Nucleotide sequenc
	10	257.4	48.2	4767	21	AAA14667	Nucleotide sequenc
	11	257.4	48.2	4818	21	AAA14669	Nucleotide sequenc
	12	244	45.7	4466	21	AAA14663	Nucleotide sequenc
	13	244	45.7	4478	21	AAA14661	Nucleotide sequenc
	14	244	45.7	4547	21	AAA14664	Nucleotide sequenc
	15	244	45.7	4571	21	AAA14662	Nucleotide sequenc
	16	244	45.7	77536	21	AAA14651	Nucleotide sequenc
c	17	242.4	45.4	77536	21	AAA14651	Nucleotide sequenc
	18	226.8	42.5	20394	22	AAF24892	Pimaricin biosynth
	19	225.2	42.2	27541	22	AAD17185	Streptomyces nours
	20	223.6	41.9	13842	21	AAZ87297	S. venezuelae macr
	21	223.6	41.9	36778	21	AAZ87318	S. venezuelae pik
	22	223.6	41.9	37948	21	AAZ87285	S. venezuelae pik
	23	223.6	41.9	38506	21	AAA75633	Nucleotide sequenc
	24	223.6	41.9	38506	21	AAZ56001	Recombinant cosmid
	25	217.4	40.7	31422	21	AAA92302	S. avermitilis ave
	26	217.4	40.7	31422	22	AAH79278	Streptomyces averm
	27	214	40.1	4689	21	AAZ87299	S. venezuelae macr
	28	210.8	39.5	43280	18	AAT80413	Tylactone synthase
	29	207.6	38.9	49377	19	AAV05287	The soraphen biosy
	30	201.2	37.7	47981	22	AAF30757	Micromonospora meg
	31	196.4	36.8	15872	18	AAT68715	Streptomyces venez
	32	196.4	36.8	15872	21	AAZ87283	S. venezuelae vep
	33	196.4	36.8	28598	17	AAT06769	Sorangium cellulos
	34	196.4	36.8	28958	18	AAT89956	Sorangium cellulos
	35	196.4	36.8	28958	21	AAA75299	DNA sequence of So
	36	191.6	35.9	44377	18	AAT78508	Platenolide syntha
	37	191.6	35.9	44377	18	AAT80414	Platenolide syntha
	38	188	35.2	53789	19	AAV21187	Amycolatopsis medi
	39	185.2	34.7	11220	21	AAZ87298	S. venezuelae macr
	40	180.4	33.8	13987	18	AAT80415	Hybrid srmG/tylG O
	41	177.2	33.2	29879	14	AAQ46806	eryA region of S.
	42	175.2	32.8	477	21	AAA58913	DNA encoding ketor
c	43	175.2	32.8	477	21	AAA58922	DNA encoding ketor
	44	154.6	29.0	50937	21	AAA09469	Streptococcus olea
	45	150	28.1	5676	19	AAV21186	Amycolatopsis medi

ALIGNMENTS

RESULT 1
AAH79279
ID AAH79279 standard; DNA; 11916 BP.
XX
AC AAH79279;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis coding sequence derivative SEQ ID NO: 3.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..11916
FT /*tag= a
FT /product= "AAG65268"
FT /partial
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX

THIS PAGE BLANK (USPTO)

```
                     TDPAHEVLVVTMGDRFGPHGAVGIILLEKKPSTWHLKLLATSCRVVSFGAGATILNWL
                     TDQGARAGAHLVADFRRTDRNRMMEIAYRFAGFADSDCPCVSEVAGASAAGVERLHLE
                     PSARPAPTTLTLTAADIAPVTVSAAG"
     gene            complement(9895. .10995)
                     /gene="fkbI"
     CDS             complement(9895. .10995)
                     /gene="fkbI"
                     /note="acyl CoA dehydrogenase"
                     /codon_start=1
                     /transl_table=11
                     /product="FkbI"
                     /protein_id="AAF86388.1"
                     /db_xref="GI:9280390"
                     /translation="MPERDALLTDLVGDRAAEWDTSGELPRDLLVRLGADGLLCAEVA
                     AEHGGLGLGSRENGEFTAHVGSLCSSLRSVMTSQGMAAWTVQRLGDAGQRATFLKELT
                     SGKLAAVGFSERQAGSDLSAMRTRVRLDGDTAVVDGHKVWTTAAAYADHLVVFGLQED
                     GSGAVVVVPADTPGVRVERVPKPSGCRAAGHADLHLDQVRVPAGAVLAGSGASLPMLV
                     AASLAYGRKSVAWGCVGILRACRTAAVAHARTREQFGRPLGDHQLVAGHIADLWTAEQ
                     IAARVCEYASDHWDEGSPEMVPATILAKHVAAERAAAGAATAAQVLASAGAREGHVVE
                     RAYRDAKLMEIIEGSSEMCRVMLAQHALALPA"
     gene            complement(10988. .11248)
                     /gene="fkbJ"
     CDS             complement(10988. .11248)
                     /gene="fkbJ"
                     /note="ACP"
                     /codon_start=1
                     /transl_table=11
                     /product="FkbJ"
                     /protein_id="AAF86389.1"
                     /db_xref="GI:9280391"
                     /translation="MTAPVKDDPVAHALVAFLKTKTRSDWPVDRDLFAEGGLTSLFAM
                     ELVVYLEKTFDVTIAGPDLQLANFRTVESMVALVHRLRAVDA"
     gene            complement(11245. .12093)
                     /gene="fkbK"
     CDS             complement(11245. .12093)
                     /gene="fkbK"
                     /note="acyl CoA dehydrogenase"
                     /codon_start=1
                     /transl_table=11
                     /product="FkbK"
                     /protein_id="AAF86390.1"

  Query Match             45.7%;  Score 244;  DB 1;  Length 77534;
  Best Local Similarity   67.7%;  Pred. No. 3.9e-22;
  Matches  359;  Conservative    0;  Mismatches  165;  Indels    6;  Gaps    1;

Qy          5 CCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCA 64
              ||  || || ||||  |  | ||  | ||  | ||  | ||| |   ||| | ||| ||
Db      61511 CCCTCCACACCACCACCCCACCCACCACCACCCCCCTCAACCCCGAACACGCCATCATCA 61570

Qy         65 CCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACG 124
               |  |    | ||     ||||  |  || |  |      |  ||   ||  |||||||
Db      61571 TCACCGGCGGCTCCGGCACCCTCGCCGGCATCCTCGCCCGCCACCTGAACCACCCCCACA 61630

Qy        125 CACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCG 184
              |  | | |||| ||  |   |  |  |   | | || || |  ||||  |  | ||||||
Db      61631 CCTACCTCCTCTCCCGCACCCCACCCCCCGACGCCACCCCCGGCACCCACCTCCCCTGCG 61690

Qy        185 ACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCC 244
              ||  | || ||||  ||||||||        |||||  |  |||||||       || ||||
Db      61691 ACGTCGGCGACCCCCACCAACTCGCCACCACCCTCACCCACATCCCC------CAACCCC 61744

Qy        245 TCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCC 304
              ||||| || || |||||||||| | || |||||||||| || ||||  .|  ||||||||
Db      61745 TCACCGCCATCTTCCACACCGCCGCCACCCTCGACGACGGCATCCTCCACGCCCTCACCC 61804

Qy        305 CCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAAC 364
              ||  ||  |||| || |||||||| | |||||||| ||  ||||    | ||| ||||| |
Db      61805 CCGACCGCCTCACCACCGTCCTCCACCCCAAAGCCAACGCCGCCTGGCACCTGCACCACC 61864

Qy        365 TCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCG 424
              |||||||| ||   |||||||||  |||||||||||||||||  ||||||||||  | |||
Db      61865 TCACCCAAAACCAACCCCTCACCCACTTCGTCCTCTACTCCAGCGCCGCCGCCGTCCTCG 61924

Qy        425 GCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACC 484
              ||   ||||| ||||  |||||||||| |||||||||| |||||||||||||||  ||
Db      61925 GCAGCCCCGGACAAGGAAACTACGCCGCCGCCAACGCCTTCCTCGACGCCCTCGCCACCC 61984

Qy        485 ACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              ||||||||||||  ||  |||||||||  |||||||||||||||  |||
Db      61985 ACCGCCACACCCTCGGCCAACCCGCCACCTCCATCGCCTGGGGCATGTGG 62034


Search completed: June 17, 2003, 17:55:34
Job time : 1038.31 secs
```

```
          ascomyceticus (ATCC 14891) contains genes for biosynthesis of
          unusual polyketide extender units
  JOURNAL Gene 251 (1), 81-90 (2000)
  MEDLINE 20323220
   PUBMED 10863099
REFERENCE 2  (bases 1 to 77534)
  AUTHORS Reeves,C.D., Wu,K., Chung,L., Revill,P. and Katz,L.
  TITLE   Direct Submission
  JOURNAL Submitted (16-FEB-2000) Kosan Biosciences Inc, 3832 Bay Center
          Place, Hayward, CA 94545, USA
FEATURES             Location/Qualifiers
     source          1. .77534
                     /organism="Streptomyces hygroscopicus var. ascomyceticus"
                     /db_xref="taxon:132248"
     gene            complement(412. .1836)
                     /gene="fkbW"
     CDS             complement(412. .1836)
                     /gene="fkbW"
                     /codon_start=1
                     /transl_table=11
                     /product="FkbW"
                     /protein_id="AAF86380.1"
                     /db_xref="GI:9280382"
                     /translation="MVLPLGVGRAPRRRTGLGALTLAAVAVLLPSPATARPAPDAPDG
                     HCARLSQLRVPAAARQQADCLDELTTAGTAATGHTDPADWAGLTPKNLATPTGVPGIQ
                     IDGYFPDTSTTNTNHGWNHDAQFVIRLPDRWNGGLVVAGTPGNREQYANDRAIADWVL
                     ARGYAYAATDKGNTGPAFYRDGSRPGDAVAEWNTRLTQLTRAARASVAQRYHRPLERT
                     LVTGMSNGGYLVRWQLENHPELYDGGVDWEGTLWRAEGPNLFTFLPAALRDYPTYAAG
                     GPAAGQAREALLRAGFPADSEFLWPYHHEVYWDLTQRVYREEFDPGFDGEKEAGTPYC
                     APGTVPGCDADYDYATRSSGVHAAVERIALTGRIGKPLITIQALSTCCCLPPGLRRVR
                     AYVRQRPGGRYRYYRIEDGTHTDALVDTFPDRLRPLLPCHRSAFTALESWLTTGHRPP
                     ANHTVVRPAGADQASLLAECRLDG"
     gene            complement(2020. .3579)
                     /gene="fkbU"
     CDS             complement(2020. .3579)
                     /gene="fkbU"
                     /codon_start=1
                     /transl_table=11
                     /product="polyhydroxybutyrate depolymerase"
                     /protein_id="AAF86381.1"
                     /db_xref="GI:9280383"
                     /translation="MQPPPFRGILTPLFPLSSSPPVGSLSRPGRRGVLTRLVAVVALV
                     LGAALLGPAPTAHAAAGLAKPGLTKADLTEVADFGTNPGRLNMYVYRPASLPAEPAVV
                     FALHGCTQDAQGYADNSGLLSFADRYGFLLVFAETTSSNNANRCFNWFQSSDNRRGQG
                     EAASIRQMAAHTVSAYGADPQRTYITGLSAGGAMTSVMLATYPDVFQAGAVVAGLPFG
                     CATDVSSAYLCMNPGTDLTADQWARRVRDGYPSWSGPWPRVAIWHGDKDTTVAPRNAD
                     ELRDQWTAVHGVSQTPDRTSVIGPNSTRHEEYLAADGSVAVEVNRVPGIGHGTPVDPG
                     TGAQQCGSTGAAYFLDSICSSYWITQFFGLSGSASDPGSLPAPSGLAATGATDTTISL
                     TWKPVDGATDYAVHRDGAQITTSATTSYTDTGLRAGTSHTYAVAARDADGKAGPLSGA
                     VTAQTTGATAVCWTAGNYAHVQAGRATTSAGYTYAKGSGQNMGLYNTFVTTTLKESPT
                     GYFTVANDTCP"
     gene            complement(3969. .4496)
                     /gene="fkbR2"
     CDS             complement(3969. .4496)
                     /gene="fkbR2"
                     /note="transcription regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="FkbR2"
                     /protein_id="AAF86382.1"
                     /db_xref="GI:9280384"
                     /translation="MVEGWQGRYFEDFTTGDVYRHPLGRTVLETDNSWLTLLTQNTAP
                     LHFDRHYSAGTTWGKPLVDSTFTLALVTGQSVTDISQHVMANLGWDRVRLPNPVFEGD
                     TIYSQSEVLNTRESASRPDVGIISVRTIGYNQDGVIVITFERTLMVYRRGHGPRFAAV
                     EPVWDERSRRAAGLH"
     gene            complement(4595. .5488)
                     /gene="fkbR1"
     CDS             complement(4595. .5488)
                     /gene="fkbR1"
                     /note="transcription regulator"
                     /codon_start=1
                     /transl_table=11
                     /product="FkbR1"
                     /protein_id="AAF86383.1"
                     /db_xref="GI:9280385"
                     /translation="MEIRELQWFTALAEHEHVTITAERLNISQPTLSRAIRRLERRVG
                     VPLFDRHQNRLRLNKYGEVFRAHALRAISEVSGAEQRITALIDPDTGTVALGFLHSYG
                     TWLVPDLLAGYRRLAPGTTFELRGDAADAVVDDVRNSRLDLGLTSPRPAGDDVEWTPL
                     RDEALCLLVPPGHRLARRRRVRTAELADEAFVALEPVFGLRQITDRLCAAAGFVPRVV
                     LESTELSTLRALVAAGLGVAVAPLTGSPSGPGDHAVAVPLDDPQARRTVGVVTLAGGP
                     RAPVVARFYDYVRSRAHAPHV"
     gene            5601. .6818
                     /gene="fkbE"
     CDS             5601. .6818
                     /gene="fkbE"
                     /note="L-carnitine dehydratase"
                     /codon_start=1
                     /transl_table=11
                     /product="FkbE"
                     /protein_id="AAF86384.1"
                     /db_xref="GI:9280386"
                     /translation="MSEPHPRPEQERPAGPLSGLLVVSLEQAVAAPPATRHLADLGAR
                     VIKIERPGSGDLARGYDRTVRGMSSHFVWLNRGKESVQLDVRSPEGNRHLHALVDRAD
                     VLVQNLAPGAAGRLASATRSSRGATEADHLRTYPATAVPAATADRKAYDLLVQCEAGL
                     VSITGTPETPSKVGLSIADICAGMYAYSGILTALLKRARTGRGSQLEVSMLEALGEWM
                     GYAEYYTRYGGTAPARAGASHATIAPYGPFTTRDGQTINLGLQNEREWASFCGVVLQR
                     PGLCDDPRFSGNADRVAHRTELDALVSEVTGTLTGEELVARLEEASIAYARQRTVREF
                     SEHPQLRDRGRWAPFDSPVGALEGLIPPVTFHGEHPRRLGRVPELGEHTESVLAWLAA
                     PHSADREEAGHAE"
     gene            6808. .8052
                     /gene="fkbF"
     CDS             6808. .8052
                     /gene="fkbF"
                     /note="similar to Rhizobium trifolii MatC; dicarboxylate
                     carrier"
                     /codon_start=1
                     /transl_table=11
                     /product="FkbF"
                     /protein_id="AAF86385.1"
                     /db_xref="GI:9280387"
                     /translation="MPNELTGVLILAAVFLLAGVRGLNMGLLALVATFLLGVVALDRT
                     PDEVLAGFPASMFLVLVAVTFLFGIARVNGTVDWLVRVAVRAVGARVGAVPWVLFGLA
                     ALLCATGAASPAAVAIVAPISVAFAVRHRIDPLYAGLMAVNGAAAGSFAPSGILGGIV
                     HSALEKNHLPVSGGLLFAGTFAFNLAVAAVSWLVLGRRRLEPHDLDEDTDPTEGDPAS
                     RPGAEHVMTLTAMAALVLGTTVLSLDTGFLALTLAALLALLFPRTSQQATKEIAWPVV
                     LLVCGIVTYVALLQELGIVDSLGKMIAAIGTPLLAALVICYVGGVVSAFASTTGILGA
                     LMPLSEPFLKSGAIGTTGMVMALAAAATVVDASPFSTNGALVVANAPERLRPGVYQGL
                     LWWGAGVCALAPAAAWAAFVVA"
     gene            8156. .8824
                     /gene="fkbG"
     CDS             8156. .8824
                     /gene="fkbG"
                     /note="O-methyl transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="FkbG"
                     /protein_id="AAF86386.1"
                     /db_xref="GI:9280388"
                     /translation="MANQITLSDTLLAYVRKVSLRDDEVLSRLRAQTAELPGGGVLPV
                     QAEEGQFLEFLVRLTGARQVLEIGTYTGYSTLCLARGLAPGGRVVTCDVMPKWPEVGE
                     RYWEEAGVADRIDVRIGDARTVLTGLLDEAGAGPESFDMVFIDADKAGYPAYYEAALP
                     LVRRGGLIVVDNTLFFGRVADEAVQDPDTVAVRELNAALRDDDRVDLAMLTTADGVTL
                     LRKR"
     gene            complement(8802. .9884)
                     /gene="fkbH"
     CDS             complement(8802. .9884)
                     /gene="fkbH"
                     /codon_start=1
                     /transl_table=11
                     /product="FkbH"
                     /protein_id="AAF86387.1"
                     /db_xref="GI:9280389"
                     /translation="MTIVKCLVWDLDNTLWRGTVLEDDEVVLTDEIREVITTLDDRGI
                     LQAVASKNDHDLAWERLERLGVAEYFVLARIGWGPKSQSVREIATELNFAPTTIAFID
                     DQPAERAEVAFHLPEVRCYPAEQAATLLSLPEFSPPVSTVDSRRRRLMYQAGFARDQA
                     REAYSGPDEDFLRSLDLSMTIAPAGEEELSRVEELTLRTSQMNATGVHYSDADLRACF
```

```
AFLELAVRAGDEVGCDRVDELTLAAPLVLPEQGGVQVQLWIGNPDASGRRSVTVYGRP
DADEDAPWTSHATGVLSASRTTSDFDATVWPPADAETLPVDGLYERLAEGGFGYGPLF
QGLRAAWRRGDEVFAEVVLPESGHTDAESFGLHPALLDSALHAASFVDLDERAAGGLP
FSWEGVSLHASGATTLRVRLAPAAGDAVAIAVADDSGQLVLSADSLILRAVAAREIDA
AAALVRDALFRLDWVPVTAVAASGTAAALVGEDPFGLRALPQFGDLAVHPDLADLAAA
DGAVPDTVLLPLTGTGPDADPVTAAHRAATEALAAVRTWLEQDERFAASRLALVTRGA
TTGHDPAAAAVWGLVRSAQSENPGRFLLVDLDADQDTPALPAAALTSEEPQLAVRGEE
LRAARLVRRPASTAEAVPAFGGEGAVLVTGGTGGLGAVLARHLVAEHGVRELVLVSRR
GGAAAGAAELVAELAESGARATVVACDVTDRAAVAELVAAHPVSAVVHSAGVLDDGMV
GTLTPERLTTVLRPKVDAAWNLHEATRDLDLKAFVLFSSVAGVLGSPGQANYAAGNAF
LDALAAHRRAAGLPGLSLAWGPWEQTGGMTGGISEDDLRRMARAGTPALTVEQGLALL
DAALDGDDAALAPVRLDLSVLRAQGEVPPLLRSLIRGRSRRAAVAGSATAGGLAQRLA
RLDAESRDELVLDLVRGQVALVLGHATGAEIDAGRAFRELGFDSLTAVELRNRLNTVT
GLRLPATLVFDYPTVSHLASYVLDELLGTEVEAEVVQRGTAAVADDPIVIVGMACRYP
GGVTSPEDLWRLVTEGTDAVSGFPVNRGWDVENLYHPDPDHPGTAYTRSGGFLHEAGE
FDPGFFGMSPREALATDSQQRLLLEASWEAIERAGIDPVGLRGSATGVFAGVMYSDYS
AMLGSPEFEGFQGSGSSPSLASGRVSYTLGLEGPAVTVDTACSSSLVAMHWAMQALRS
GEISLALAGGVTVMSTPAVFVDFARQRGLSPDGRCKAFSDSADGVGWSEGVGMLVLER
QSDAIRNGHQILAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALASGGLSAGDVDVV
EAHGTGTTLGDPIEAQALLATYGRDRDPEQPLLLGSVKSNIGHTQAAAGVAGVIKMVM
SMRHGVLPRTLHVDAPSSHVDWTEGAVELLTEQTAWPETGRPRRAAVSSFGISGTNVH
TVLEQAPGTTVPAPAAPERTAGAVPLLLSGRTRDALRAQAARLLTHLQNHPEPSLADL
GHSLATTRSRFERRAAVIAQDREGLLASLGSLAAGRPDPAVVEGEAAGRARVAVMFSG
QGSQRAAMGRELYETQPRFAAAFDEVCAALDPLLDRPLREVVFAAEGSEEAALLDRTG
WTQPALFAVEVALYRLVESWGVRADFVTGHSIGEIAAAHIAGVFTLQDAARLVAARAT
LMEALPSGGAMVAVQATEEEVAPLLGEGLSVAAVNGPTSVVVSGDEDPAVELAAEFSG
RGRRTKRLRVSHAFHSPHMDAMLDAFRTVAETLSYAAPRIPLVSDLTGRRADDAEVRT
ADYWVRHVREAVRFADCVRTLRDAGATLFLELGPDGLLTAMAEDTLGDERYDHNTALV
PLLRADRPEESAAATAAARLQIHGVDLDWTAYLAGTGARRVDLPTYAFQHAHYWPQLP
SAAPSPAGDPADQKLWAAVERGDAAELAAVLGLDEDSLTPLDSLLPALSSWRRGNQEK
ALLDTLRYRVEWTRLSKPAAPVLDGTWLLVSSDATADDETELLDGLAEALGAHGARVR
RLVLDADCADRAVLGARLADTENADNTAQVLSVLPLDERPTDGPAGFTQGLALTIALV
QALADTGAHGRLWTATRGAVSTGPADPVTHPAQATAWGMGRGVALEHPRLWGGLVDLP
ADFDRGAGQRLAEVLAVKDAPDGEDQVALRATGVHGRRLVRHIVDELPSADQFTASGS
VLITGGTGGLGAETARWLARSGAAHLVLTSRRGPDAPGAAELRAELEQSGASVSIVAC
DVADRDALAAVLDGLSADQPLTGVVHTAGVGHYGPLDALTPAEFAGLTAAKLAGAAHL
DNLLGDRELDFFILFGSIAGVWGSGDQSAYGAANAYLDALALARRARGLAATSIAWGP
WGGTGMAADDAVSGTLRRQGLGLLDPAPALTEMRRAVVRQDVTVTVADVDWTRYAPLF
TSARPSALISDLPEVRALAAENTPADTGDASEIVQRVRSLSEPEQLRLLTDLVRTEAA
TVLGHSSAGAVPEDRAFREIGFDSLTAVELRKHLGAATGLSLPSTMVFDYPTPLELAQ
YLRAEMVGSVLEVAGPVATGGTDDEPIAIIGMSCRYPGGVSSPEQLWDLVLSGTDAIT
DFPVNRGWNTAGLYDPDPDHPGTTYSTQGGFLHEADEFDPMFFGISPREALVMDPQQR
LLLETTWEAFERAGLTPDTLRGSLTGTFIGSSYQEYGMGAGDGAEGHLVTGTSPSVLS
GRLAYVFGLEGPAVTVDTACSSSLVALHLACQALRNGESNLAVAGGATVMTTPNAFVA
FSRQRALAQDGRCKAFSESADGMTLAEGVGIVLVERLSDARRNGHPVLAVIRGSAINQ
DGASNGLSAPNGPSQQRVIRQALANARVAPGEIDLLEAHGTGTPLGDPIEAQALFATY
GRTRTPETALLLGSVKSNIGHSQSAAGVASIIKMVMALRHGVMPQTLHADEPSSHVDW
SPGTVRLLGENTDWPQTGRPRRAAVSSFGISGTNAHVILEQETEAPAAEDEQLAPAPL
PVAAGVVPWLLSARGAAALREQADRLLTHLVTADPAARPIDIGLSLATSRALFEHRAV
VVPPAGTDPLEALRAVAADGPSGVVARGVADVAGRTVFVFPGQGSQWAGMGAQLLDES
PVFAERIAECAAALAEFTDWNLIDVLRGAEGAPTLERVDVVQPASFAVMVSLAAVWRA
QGVEPDAVVGHSQGEIAAAVVSGALSLRDGARVVTLRAQAIGRSLAGRGGMMSVALPV
AEVEARLEAFEGRVSVAAENGPRSSVVAGEPEALDELHAQLTAEEIRARRVAVDYASH
SPHVEDLHDEILELLAEVAPRTSEIPFFSTVTGDWLDTTVMDAGYWYRSLRGRVLFAD
AVRDLIAADHRAFIEVSSHPVLAMSVQDMIDDAGVAGVASGTLRRDNGGLDRFLLSAA
EVFVRGVQVDWAAVFEGTGASRVDLPTYAFQHENLWAMAAAPEAVTAADPEDAAFWTA
VEDGDVSALTAALGTDEDSVAAVLPALSSWRRARKERSTVDSWRYRPTWKPVTKLPQR
TLDGTWLLVSADGVDDTDVAEALETGGAEVRRLVLDESCTDRAVLRERLTDADGLTGI
VSVLAGAERTGAVPGTGLVLGVALTVALVQALGDAGIDTPLWALTRGAVSTGRSDKVT
APVQAQVTGIGWTAALECPGRWGGVVDLPETLDARAGQRLAAVLAGALGDDDQIALRS
SGVFTRRIVRADAAPDGSARDWKPRGTTLVTGGSGTLAPHLARWLAEQGAEHLVLVSR
RGPEAPGAAELRAELAERGTETTLAACDITDRDAVAALLESLKAEGRTVRTVVHTAAT
IELHTLDATTLDDFDRVLAAKVTGAQILDELLDDEELDDFVLYSSTAGMWGSGAHAAY
VAGNAYLAAALAEHRRARGLTALSLSWGIWADDLQLGRVDPQMIRRSGLEFMDPQLALS
GLKRALDDDEQVIAVADVDWETYHPVYTSARPTPLFDEVPEVQRLTAAAEQSAGDPAR
GEFAAALLALPAAEQHRKLLETVRTEAASVLGLSSAEDLTDQRAFRDVGFDSLTAVGL
RNRLASVTGLTLPSTMVFDYPNPAALAGFLHSELADVHSAGAVAVTAGAPVDDDPIAI
VGMSCRYPGGITSAEQLWRVSLEEVDAVSVFPADRGWDAEALYDPDPDASGRTYSVQG
GFLRDVADFDPGFFGISPREALSMDPQQRLLLETAWEVFENAGLDPVAQRGSRTGTFI
GASYQDYGAAVPGSEGSEGHMITGSLPSVLSGRVSYLFGLEGPAVTLDTACSSSLVAI
HLACQSLRNGESTLALAGGASIMSTPMSFIGFSRQRALAEDGRCKAYAEGADGMTLAE
GVGLILLERLSDARRNGHEVLAVIRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAG
VEANDIDVLEGHGTGTALGDPIEAQALFATYGKDRDPERPVLLGSVKSNIGHTQMASG
```

```
VASIIKLVHALREGVAPKSLHIDQPSTHVDWSSGTIQLLTERTEWPETGRPRRAAVSS
FGLSGTNVHTVLEQAPAADAPAAEDTPAPRDALVPVLVSGRGEAALRAQAGALLDLLA
ERPGIHPTDLAFSLATSRAALEHRAAVVADDHEALVRGLTALRDGLPGAGLVQGRTGR
GRTAFLFTGQGSQRLGMGRELYERHPVFADALDAVLARIDGTTERPLRDVLFAAEGSQ
DAALLHRTGYAQPALFALEVALFRLLESWGVTPDYLAGHSVGEIAAAHVAGVLDLDDA
CTLVAARGRLMQALPEGGAMVALEAAEDEVLPHLEGLADQVSVAAVNGPRSVVVAGEE
EPVLALAAHFAEQGRRTKRLRVSHAFHSPLMDPMLDDFAAVARALTYHAPSIPFVSNV
TGTLAAPEQVCTADYWVSHVRSAVRFADGIGWLSTQGGVQTFLELGPDGVLSGMARES
LTDASRTALLPTLRGDRPEEQALVTAVAAAHAHGFDVDWTAWFQGSGARRVALPTYAF
QRERYWPDTTAAGITAPAPGSALDAEFWAAVEHADVASLTASLGLDDATVTAMVPALT
AWRQRRGEQSALDSWRYRVTWKPRGGAPGAAPTGRWLVLVPAEHRDEATAAWAADVEA
ALATATVRVEVTGTDRAALAARLTEAADGDTFQGVLSLLALAPGDAGHPGAPAALTLT
ATALQALGDARIDAPLWNITRGAVAVGRSEQVTAPEQAAVWGLFRAAAALELPARVGGS
VDLPEDLDTQAARRLRGILAAADGEDAVAVRASGVFLRRLAHSPAADTVGSAFDPAAG
TVLITGGTGGIGGHLARRLARDGAAHLLLTSRRGPDAPGAGELRAELEESGARVTIAA
CDAADRDALAALLATVPEDAPLTAVFHTAGVVDDHVVDELTPESFATVLHAKTVAARH
```

Query Match 45.8%; Score 244.4; DB 1; Length 113193;
Best Local Similarity 66.1%; Pred. No. 3.2e-22;
Matches 353; Conservative 0; Mismatches 181; Indels 0; Gaps 0;

```
Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  ||| |||||||||| |||||||  |||  |   |  |||| ||| ||||||||
Db      84111 GGCACCGTCCTGATCACCGGCGGCACCGGCGGACTCGGCGGCGTCCTCGCCCGCCACCTC 84170

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
                |||| | |    | ||  |  || || ||| |||||||   ||||||   | ||  |
Db      84171 GTCACCGAACGCGGCGTCCGCCGGCTGCTGCTCGCCAGCCGCCGCGGCCCGGCCGCCGAC 84230

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
               ||| |    |||| |  | |||||      |  |||  |    ||  | || || ||
Db      84231 GGCGCCCCCGCCCTCGTCGACGAACTCACCGCACTCGGCGCCTCGGTCGACGTCGCCGCC 84290

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              |||||| || | ||| ||   |||      |||||| |    | ||| |    |||
Db      84291 TGCGACGTCACCGACCGGGAGGCGCTCGCCGCCCTCCTCGCCCGGGTGCCCGCCGCGCAC 84350

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              || || | | ||| | |||||||| ||| |  |||||||| |  | ||  |  ||||
Db      84351 CCGCTGAGCGCGGTCGTGCACACCGCCGGCGTGGTCGACGACGGCGTCATCGGCTCCCTC 84410

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              ||| ||   |  ||| ||| ||| || ||  |||| ||| ||  ||||   |  || |||
Db      84411 ACCGCCGAGCGCCTCGACACCGTGCTGCGGCCCAAGGCCGACGCCGCCTGGCATCTGCAC 84470

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              | |||||||   ||   ||| |||| |||| |||| |||| |||||||| ||||||||
Db      84471 GAGCTCACCCGGGACCTGGACCTGGACGCGTTCGTGCTCTTCTCGTCCGTCGCCGCCACC 84530

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              |||||   |||||  || | |||||||| || | |||||||| ||| ||||||||||||
Db      84531 CTCGGCAGCCCCGGACAGGGCAACTACGCGGCCGGCAACGCCTTCCTGGACGCCCTCGCC 84590

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              |||||||||| ||| |  ||| |||||||||  | ||||||||||| |||||
Db      84591 GCCCACCGCCACGCCCTCGGCCTGCCCGCCACCTCCCTCGCCTGGGGCCCCTGG 84644
```

```
RESULT 15
AF235504
LOCUS       AF235504               77534 bp    DNA     linear   BCT 19-JUL-2000
DEFINITION  Streptomyces hygroscopicus var. ascomyceticus FK520 biosynthetic
            gene cluster, partial sequence.
ACCESSION   AF235504
VERSION     AF235504.1  GI:9280381
KEYWORDS    .
SOURCE      Streptomyces hygroscopicus var. ascomyceticus.
  ORGANISM  Streptomyces hygroscopicus var. ascomyceticus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 77534)
  AUTHORS   Wu,K., Chung,L., Revill,W.P., Katz,L. and Reeves,C.D.
  TITLE     The FK520 gene cluster of Streptomyces hygroscopicus var.
```

```
Qy        61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
             ||||  |||   | |||   ||||| ||||||||||||||    || ||| || |||||
Db     20052 GTCACCACCCACGGCGCCCGCGACCTCATCCTCACCAGCCGCCAAGGACCCAACGCCCCC 20111

Qy       121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
             |||     | ||||||| ||||.|||||||| | |||  ||   ||  || || | ||
Db     20112 GGCGCCACCGAACTCACCAACCAACTCCAACAACACGGCGCCCGAATCCGCACCATCTCC 20171

Qy       181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
             ||||||  || | |||    |   |||     |||| || ||   |||             |
Db     20172 TGCGACCTCACCGACCGCACCGCCCTCACCACACTCATCGACGAGATC---------GGC 20222

Qy       241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
             |||||||||  |||| | |||||||| ||    ||||| |||| ||||| ||  | |||||
Db     20223 CCCCTCACCGGCGTCGTGCACACCGCCGGAGCCCTCGCCGACACCACCATCGACCACCTC 20282

Qy       301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
             | ||||  |  ||||| || |  |||  | |||||||| ||  ||||       ||||||
Db     20283 AACCCCGACACACTCACCACCACCCTCACCCCCAAAGCCAACCCCGCCTGGTGGCTCCAC 20342

Qy       361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
             ||||||||||| ||  |  |||  |   |||||| ||| ||| || | |||||
Db     20343 GAACTCACCCAAGACCACGACCTGGCGCTGTTCGTCGTCTTCTCATCGGTCGCCGGTGTG 20402

Qy       421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
             | |||    | || || |||||||||| || ||||||  || ||||||||||    |
Db     20403 CTTGGCAACGCGGGGCAGGCCAACTACGCGGCCGCCAACAGCTTCCTCGACGCCCTGATC 20462

Qy       481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
             ||||||||  |  |||   ||| |||| ||||  | ||||||||||   |||
Db     20463 ACCCACCGCCGCCGCCAGGGCCTGCCCGGCACCTCCCTCGCCTGGGGCTGGTGG 20516
```

RESULT 14
AF357202

```
LOCUS       AF357202               113193 bp    DNA     linear   BCT 17-JUL-2001
DEFINITION  Streptomyces nodosus amphotericin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF357202
VERSION     AF357202.1  GI:14794889
KEYWORDS    .
SOURCE      Streptomyces nodosus.
  ORGANISM  Streptomyces nodosus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 113193)
  AUTHORS   Caffrey,P., Lynch,S.V., Flood,E.M., Finnan,S.M. and Oliynyk,M.
  TITLE     The amphotericin biosynthetic gene cluster from Streptomyces
            nodosus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 113193)
  AUTHORS   Caffrey,P., Lynch,S.V., Flood,E.M., Finnan,S.M. and Oliynyk,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2001) Industrial Microbiology, University College
            Dublin, Belfield, Dublin, Ireland
FEATURES             Location/Qualifiers
     source          1. .113193
                     /organism="Streptomyces nodosus"
                     /db_xref="taxon:40318"
     gene            complement(4. .1824)
                     /gene="amphG"
     CDS             complement(4. .1824)
                     /gene="amphG"
                     /note="possible amphotericin export protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphG"
                     /protein_id="AAK73498.1"
                     /db_xref="GI:14794890"
                     /translation="MAPDVPEEHEEERESEQPVRRLAALLRPHRRSVGLALTAGVVGI
                     LLNAFGPLLLGRVTDLIADGVLGHGGPAPGVDFGALGRLLMILLVLYVVASVFMLVQN
                     WLVASVVRLLIHDLRHRAQEKLARLPLRYFDRKPAGETLSRGTDDVDNLQQTLQQTLT
                     DLISSVFSLVIMLSLMLIISPSLAGVMLLSVPVSGLLAAWISKRAQPQYAAQWSASGK
```

```
                     LTAHVEEMCAGHALVKAFDRRAEAEQRFDERNEAVYRAGSGAQFASGAIEPVMMFVAN
                     LGYVAVAVVGAWKVVNGSLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGVASAQRV
                     FTLLDAPEQEPEPDRPLAVERVEGRVEFQDVSFRYSPDTPLIEGLSLSVEPGSTVAVV
                     GPSGAGKTTVANLLMRFYEIDSGRILLDGTDTAAMNRDDLRSRFGLVLQDTWLFKGTI
                     AENIAYGSPGATRADIVEAARATYADRFIRTLSQGYDTVLDDESGGVSAGEKQLITVA
                     RAFLARPAVLVLDEATSSVDTRTELLIQRAMNTLRPGRTSFVIAHRLSTIRDADVIVV
                     MKSGRIVEQGTHDQLIDAQGAYARLHAARADAPAADVTVG"
gene                 complement(1805. .3628)
                     /gene="amphH"
CDS                  complement(1805. .3628)
                     /gene="amphH"
                     /note="possible amphotericin export protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphH"
                     /protein_id="AAK73499.1"
                     /db_xref="GI:14794891"
                     /translation="MAPSVRWPTPFQRAAAGRRQPGCHSVLLRLMRTQLRPYAGSVVA
                     LVVLHLVQILGTLLLPTLGAALIDEGVVRHDSDRIGTIGATMAVVALVQIAAALGAAA
                     LGARTSTALGRDLRSAVFRRVLDFSAREIGRFGTPSLLTRSVNDVQQVQNLAQSGLGI
                     FVAAPLMCLGSVLLALRQDVTLALILVPMVLVVAVCFGLLLSRMAALYARLQQTLDRI
                     GRLLRERITGVRVVRSFARDAHEGERFTRTNEELLGLSLGVGRLIAVMLPSVLLLMNL
                     FTLGLLWVGARRIDSGSMQIGALSAFLSYLSLILMSVVMLAFVFLNVPRARVCAERIT
                     EVLQAETDVVPPASPRPMAGPAGQVELVGAEFRYPGAENAVLRDLSLTLRPGERVAVL
                     GSTGSGKTTLLHLILRLVDVTAGEVRIGGTDVRELDPSVMAAAVGYVPQRPYLFAGTV
                     ASNMRFGRPDATDEELWEVLRIAQHEGFVTRLGGLDTEIAQGGTTVSGGQRQRLAIAR
                     ALLRRPAIYLFDDSFSALDQSTEAALRKALVPYTEGATVITVAQRVASVRDADRIVLL
                     DQGGIAATGTHDALLRDSPTYREIALSQRTREETAHGAGRS"
gene                 3840. .4874
                     /gene="amphDIII"
CDS                  3840. .4874
                     /gene="amphDIII"
                     /note="mycosamine biosynthesis protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphDIII"
                     /protein_id="AAK73500.1"
                     /db_xref="GI:14794892"
                     /translation="MPKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKFRV
                     SRLASELSFVDGDLMDQGSLVSAVDKVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                     VLRVLEAIRMVSGLSMSRTAGTEGQIRFYQASSSEMFGKVAETPQRETTLFHPRSPYG
                     AAKAYGHFITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKLGLQDKL
                     ALGNMDAVRDWGYAGDYVRAMHLMLQQDAPDDYVIGTGEMHTVRDAVRFAFEHVGLDW
                     KDYVVVDPDLVRPAEVEVLCADSSKAQAQLGWKPSVDFQELMRMMVDADLASVSRQNE
                     LDDLLLAHSW"
gene                 5042. .33574
                     /gene="amphI"
CDS                  5042. .33574
                     /gene="amphI"
                     /note="polyketide synthase multienzyme polypeptide housing
                     extension modules 9, 10, 11, 12, 13, and 14"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphI"
                     /protein_id="AAK73501.1"
                     /db_xref="GI:14794893"
                     /translation="MDNEQKLRDYLKLATADLRRARRRVGELESASQEPIAIVGMTCR
                     YPGGVSSPEDLWRMVEAGENGVTPFPTDRGWDLEALASSPTSRGGFLHDAPEFDADFF
                     GISPREAVAMDPQQRVVLESAWEAFERAGINPTSVKGSRTGVFIGAMAQDYRVGPADG
                     AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                     AGGVTVMSGPGTFIEMGRQGGLSVDGRCRSFGDTADGTGWAEGVGILVLERLSDAIRN
                     GREILAVVRGTAVNQDGASNGLTAPNGPSQQAVIEQALVNARLSAGDIDVVEAHGTGT
                     TLGDPVEAQALLATYGQQRDEDKPLLLGSVKSNISHTQAAAGVAGVIKMVMAMRHGVL
                     PRTLLADEPTRHVDWSQGAVRVLTENTEWPATGAPRRAAVSSFGISGTNAHTIVEQAP
                     EPEPADPEDDAPSTPAAVTGVLPVLLSGRSPEVLRAQAAALLTTLGTGTPPAPADLAY
                     SLATTRTAFEHRAVLLASDLPELTGRLTAIAEGTDPAVLADTVTGTARTETRLAVLFT
                     GQGAQRLGAGRELAARFPVFAAALDAALDAFTPHLDVPLRKVLWGEDADRLDRTEYAQ
                     PALFAVEVALYRLLESFEVKPDHLAGHSVGEIAAAHVAGVFSLDDAATLVAARGRLMQ
                     ALPEGGAMVAVQASEDEVAPLLAGHEDLVSLAAVNGPSAVVLSGDETTVTELAARLAA
                     DGRKTSRLRVSHAFHSPLMAPMLDEFRNVVEGLTLHSPLLPVVSDVTGEPATVAQLTS
                     PDYWVDHVRQAVRFADGIDWLARHDVTAFLELGPDSVLSAMAQNCLDAAGSDALTVPA
                     LREGRPEDHTFTAALAALHTQGTALHWDACFTGTGARRTDLPTYTFQRRRYWPRAVQG
                     GAADLRSVGLGAAHHPLLSAAVSLADSEGALLTGRISLLSHPWLADHTVRGATLLPGT
```

FQTAGPPVTAVLQVAGPETLPTTLTGEVELGPLQDALALDELTREAHPALFVLSAPID
NGPAHAVFQAVANRRHAAGLPALALSWGPREELTSRERLALFDAATTVDDASVLAIRH
ELAASAVDTGAAGRLARLPAAERGRALLDLVLSEVGVITGGRPVHDRQSFRDIGMTST
SAVELRGRLAAATGLALPATLAFDYPNPDALAAYLGGRLFGDTSAAGPDTEPTDPAAP
SGEPIAIVGMACRLPGGVTSPEELWELLSSGGDGISGFPTDRGWDLEHLYHPDPDHPG
TSYVREGGFLHDAGDFDAALFGISPREALAMDPQQRLLLETTWEVLERAGIDPTSLRG
SRTGVFAGVMYHDYISRLRTVPPELEGYLSSGGSGSVVSGRVSYTFGFEGPAVSVDTA
CSSSLVALHLAVQALRTGECTLAVAGGVTVMASPSSYIEFSRQRGLSPDGRCKSFAAD
ADGAAWSEGAATLLVERLSDAQRLGHPVLGVVRGTAVNQDGASNGLTAPNGPSQQRVI
RDALTNAQLTATDIDLIEAHGTGTKLGDPIEAQAVLAMYGQDRDLPVWLGSLKSNIGH
TQAAAGVAGVMKAVLSMRHGVMPKSLHIDEPTPEVDWSAGTVELLAEARDWPGADERP
RRAGVSSFGVSGTNAHVILEQAPPAPEPTESAEPVEPSEAAGPSGPVGLELGPSGPVP
WVLSAQNETALKDLARQLVTETEGTVHQIAHSLITQRSTTLPHRAVLVGSTREEFATA
LEQITRHTPDGGKTVLVFPGQGSQWVGMGQRLWETSPLYRDTIHEIDTALSEYIDWSV
TDILTSQPGAPTWDRVDVIQPVLFTTMTALATLWQHHGLTPDAIIGHSQGEIAAAYTA
GALTLQDAARITTLRSKALLPLTGHGTMISVLAPPEKIQEIADPDHWNTRIWIAAHNA
PNTTTITGDTQAITHLTQQLSHHRIMRWQLPGVDFAGHSGHIDQLHNQLQKILAPITP
QKPTIPFYSTTDNTWITDTTLDADYWYRNLRQPVLFQPAINHLTNNGYTTYIETSPHP
TLTPSIQETNPNTTTIHTLHNNQNDTHALLTALGHAHTHGHPITWHTLIPPTKTTPLP
TYPFQHTRYWLNDKPTVDTQDATQLGLIPTPHPLVPATTTLAETGATILTGHINPAHH
PWTTDHAVNGTPLLPGTALVDMALHAGDHTDHTTLDELIIHTPITLTHPTTIQLTISP
PDNTNTHTLTVHTRNSPDGPWTAHATGTLTQHQHTPDPPQTTWPPTDAEPIPLDGTYQ
RLATTGLEYGPTFQGLDALWRRDSRLFADVRLPNTDHTDSADGYGIHPALLDAALHAL
VAVSDGDEIRLPFAWTGVTLHATGATKLRVELEWDASGSASLRAFDPSGQPVVTVESL
ASRVISPEQLRSASARSGPLYRPEWIAWSGAGTPGGAVEFERYEVPTVGDDVLADTHR
RSTETLTRIQQHLAADTTTPLVVEATHDDLAGAAVWGLIRTAQTEHPGRIILIDTDNH
PTSRKALPTLIASGEPQARIRDGAITLPRLTVVAPTEPAGVGEGTVLITGGTGSLGSL
IARHLVTEHGVRDLVLVSRQGPDAPGATELIAELQQLGAQVRVRSCDTTDRAALAALV
DEVGPFAGVVHTAGALADATVDNLDADALATTIAAKAHAAWWLHEHTQDHGLTLFMVF
SSLASVLGSAGQANYAAANSFLDALATHRRTLGLPGTSLAWGSWERTSGMSAHLTRAD
HDRLARAGLRPLTDEQGLEFFDGALAVDLGQVVTATFDLPAIGRSGAVPVLLRGLVRA
PRRAAVGADSGGAEPWTERIATLAPDARERSVLTSVIEQVAQALGHPDPGAIEPDQAF
KSLGFDSLTAVEFRNRLSAMTGLQLPATLVFDYPAAGAVTRYLLDQVSPAPAAAGPDD
DPDERELRRVLERVPVARFREAGVLEALLKLAGPDPARTPDAAEPDEDGLIDAMDTEA
LIRHVMDGTGA"

CDS 9637. .14925
/note="orf2"
/codon_start=1
/transl_table=11
/product="PKS module 2"
/protein_id="AAC38062.1"
/db_xref="GI:2624949"
/translation="MTNEQLVEALRVSAKENARLRRENASLQEGAAEPVAIVGMACRL
PGGVDSPESLWELMASGGDAVSPFPTDRGWDLDGLYDPDPDHLGTAYVREGGFLDRAG
AFDPDPFGISPREALAMDPQQRLLLETTWEVLERAGIDPTTLRGSRTGVFTGIMYQDY
VTRLRDFPAELEGYLGNGSAASVASGRVSYTFGFEGPAVSLDTACSSSLVALHLAVRP
ALRRLLARVAGGVTVMSVPHPYVEFSRQRGLAFDARCKSFDASADGTVLSEGVGLLLV
ERLSDALRLGHPVLAVVRGTAVNQDGASNGLTAPTGPSQQRVIRDALTDARLTVADVD
LLEAHGTGTRLGDPIEAQALIATYGQDRDQPVWLGSLKSNIGHTQAAAGVAGVIKTVM
AMRHGVMPRTLHVSEPTPNVDWSAGEVRLLTWARDWPSTDQRPRRAAVSSFGISGTNA
HVVLEQAPAAPEPDVPVTPPAGPVPWVLSAHTGPALSALAERLAAHVEEDEAPLGDVA
HTLNTTRATLPHRAVVVGETREEFVAALRAVRPTRPRRGPLAVMFTGQGSQRPGMGRQ
LYEHFPVFAQALDEVFALATPGLREVMFDPDQAETLQRTDHAQIALFAFETALYRLWE
SWGLRPDMVCGHSVGEITAAHVSGTLTLPDAVHLVTTRGTLMQNLPPGGAMLAVATDP
HTLQPHLDNHHDTISIAAINGPHATVLSGDRTTLHHIATQLNTKTNWLNVSHAFHSPL
MQPILQPFTTTLNTLTHHPPHTPLISMLTATPTHPDTTHWTQHITAPVRYTDTLHHLH
HHGITTYLEIGPDTTLTALARTTLPTTTHLIPTTRRNHNEVRSTNEALGRVFSVGHSV
DWRALTPTGRRTSLPTYPFQRRDFWLHDAAGGGAEVEGAGLGTTDHPLLGAVTTVADT
GELILSGRLSTSTHPWLTDHTVNGTVIVPGTALIDLALHAAEHTDHTTVDELVIHTPL
ALTHPTPSQSTVGAETDTGHRPIALHSRDTTGTWTRHTTGTLSNQTHPAAELSTWSPT
DAHQIDLTTAYQQLADTGLHYGPAFQGLHTLYQRDNTLFAEIELPEAAGLPTGHILHP
ALLDAALHPAAVSAARDEESGIQLPFSWAGVTVHSTGATHLRVRLDTDDTGAIRRLAT
DTTGQPVITIDSLTTRPLHPDQLHAAQPETNHLYTLGWSPTPITPTTTGHFERYTVPT
AGDDILADTHRITAQTLTRIQQHLAADTTTPLVVEATHDDLAGARVWGLIRTAQTEHP
GRIILIDTDNHPTSRDALASVVASGKPQARIRDGAITLPRLTRLSVDTRASTERRPDL
SMGPVLVTGGTGGLGSLIARHLVTEHGVLDLVLVSRQGPDADGATELTNQLQQHGARI
RTISCDLTDRTALTTLIDEIGPLTGVVHTAGALADTTIDHLNPDTLTTTLTPKANPAW
WLHELTQDHDLALFVVFSSVAGVLGNAGQANYAAANSFLDALITHRRRQGLPGTSLAW
GWWQREGGMTAHLTQADHDRMTRAGIHGLTDAEGTTLFDTALDRGLGAVAPVKLHLPT
LNRADTVPAVLRGLVRAVRPAARQATVPAESWADRIAALMPEERDRAVLDLVRQQSAR
VLGHSDADRIDPGQAFKDTGFDSLTAVEFRNRVTTATGLRLPATLIFDHPTPNALADH
LLAQLGPLGDGGSSSPEVVRDLERFERQLFDTRPDDEVGTAIAVRLRSILAHLDAPAT
APAARTASAQVLDTLEDASQSELLSFIDKEFGRASN"

CDS 14964. .21266
/note="orf3"
/codon_start=1
/transl_table=11
/product="PKS module 3"
/protein_id="AAC38063.1"
/db_xref="GI:2624950"
/translation="MADSKTLTDYLKWVTADLYRTRERLAEVESASAEPIAIIGMACR
FPGGVTSPEALWELVSSGTDAISGFPDDRGWDMEGFFDSDPAQEGTSYMREGGFLTNA
TGFDPAFFGISPREAVAMDPQQRLLLETTWEVLERAGIDPTSLRGSRTGVFTGLMANS
YTTRVRRIPGELESYAGNGEAASVASGRVSYTFGFEGPAMSVDTACSSSLTALHLAVR
ALRSGECPLAVAGGVTVFATPRIFSDLSRQRGLARDGRCKAFSTSADGTGLSEGVGLL
LVERLSDAQRLGHSVLGVVRGTAVNQDGASNGLTAPNGPAQQRVIKAALADARLTAAD
IDLIEAHGTGTKLGDPIEAQAIIATYGQDREQPLWMGSLKSNIGHSQAAAGVAGIIKI
VMAMRHGVMPKTLHIDEPTPEVDWSAGTVELLAEARDWPGVDERPRRAGVSSFGVSGT
NAHVILEQAPPAPEPTESAEPVEPSEAAGPSGPVGLELGPSGPVPWVLSAQNETALKD
LARQLATVDGSVRDIAHSLISTRAALPHRAVVVGSTREEFATALEQITGHTPDGGKTV
LVFPGQGSQWVGMGQRLWETSPLYRDTIHEIDTALSEYIDWSVTDILTSQPGAPTWDR
VDVIQPVLFTTMTALATLWQHHGLTPDAIIGHSQGEIAAAYTAGALTLQDAARITTLR
SKALLPLTGHGTMISVLAPPQTIQDIADPDHWNTRIWIAAHNAPNTTTITGDTQAITH
LTQQLSHHRIMRWQLPGVDFAGHSGHIDQLHNQLQKILAPITPQKPTIPFYSTTDNTW
ITDTTLDADYWYRNLRQPVLFQPAINHLTNNGYTTYIETSPHPTLTPSIQETNPNTTT
IHTLHSNQNDTHALLTALGHAHTHGHPITWHTLIPPTKTTPLPTYPFQHTHFWLNEKT
ADADVESAGLSRTDHPLLGATTTLAHSGETVLTGRLSLTHHGWIADHAVNGTPLLPGT
ALVDMALHAGDHTDHTTLDELIIHTPITLTHPTTIQLTTTPTHDDNTLNLTIYTRTAP
EQPWIQHAAGTLTRANTAPSSPETIWPPTNAEPIPLDGTYQRLATTGLEYGPAFQGLH
ALWRRDSDNRLLADVRLPDTIDNTNSTDCTDGYGIHPALLDAALHALVAVSDGDEIRL
PFAWTGVTLHATGATKLRVELDIDDDGTLSLSATDTSGQPVVTVDSLTTRPIDPAQLR
ARPRQTDDLYALDWLPWSGSVAPAPEFEQYEVPKVGDDVLADTHRITAQTLTRVQQHL
AADTTTPLVVEATHDDLAGAAVWGLIRTAQTEHPGRIILIDTDNHPTSRDALASVVAS
GEPQARIRDGAITLPRLTTTAVTDALTPPADADAWHLDTSERGTLENLTLTGSDTATR
NLLPGEVRVAVRAAGLNFRDVLIALGMYPGEATMGSEAAGIVLDIGPDITDLTPGDRV
TGLFPNALANTAVADRRMITPIPASWSNTDAAAFPIVYLTAYYGLHDLAALKPGERVL
IHAGTGGVGTAAIQLAQHLGAEVFATAHPTKWPHLRAMGLDDHHIATSRTPDFETTFH
TTTQGQGIHVILNSLTDDLTDASLRLLTPNGRFIEMGKTDVRDTEEVARTHPGVTYRA
FDIADDAGPDRIGAMLVELTGLFERGALRGLPVRCWDVRESVDAFRFMAQARHIGKVV
LTIPHTPDGTILITGGTGNLGSLIAHHLVTTHGARDLILTSRQGPNAPGATELTNQLQ
QHGARIRTISCDLTDRTALTTLIDEIGPLTGVVHTAGALADTTIDHLNPDTLTTTLTP
KANPAWWLHELTQDHDLALFVVFSSVAGVLGNAGQANYAAANSFLDALITHRRRQGLP
GTSLAWGWWQREGGMTAHLTQADHDRMTRAGIHGLTDAEGTTLFDTALDRGLGAVAPV
KLHLPTIARSGSVLPVLRALVRAARPTAGQAAVQDGTWADRVAAVAPEERRAVVLELV
RAQAATVLGHSGAEHVSPTQAFKDSGFDSLTAVEFRNRVTTATGLRLPATLIFDHPTP
NTLADHLLAQLAPAYTGPSLADLDGIEVTLRTLAEDGERRPAVVARLKQWLSTLGAAQ
PAADTALDLDAASAREIFDFLDGKAG"

CDS 21363. .27176
/note="orf4"
/codon_start=1
/transl_table=11
/product="PKS module 4"
/protein_id="AAC38064.1"
/db_xref="GI:2624951"
/translation="MANEDEFLHYLKKAAADLRDARQQIQELEARDREPIAIVSMACR
LPGGVDSPESLWDLVAAGRDGVSAFPIDRGWDLEGIYDPDPDHPGTTYARDGGFLDDA
AGFDPAFFGISPREALAMDPQQRLLLETTWEVLERAGIDPAALRGSRTGVFTGVIHND
YADWSADDTDQLEGYLGNGSAASVASGRVSYTFGFEGPAVSLDTACSSSLVALHLAIQ
ALRSGECTLAITGGVTVLASVAPFIEFSRQRGLADDGRCKAFAADANGFGMAEGIGLL
LVERLSDAQRLGHPVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIRDALTNAQLTATD
IDLVEAHGTGTKLGDPIEAQAIIATYGQHRDQPVRLGSLKSNIGHAQAAAGVAGIIKT
VMAMRHGIMPRTLHIDEPTPEVDWTTGTVELLTESRDWPTTDQHPRRAAVSSFGVSGT
NAHVILEQAPTTPEPHASEAPPAGPVPWILSAHTEPALRELARRLVTRHEESVTDVAY
SLLHGRAALPCRAVMIGSTREDFAAELEALTCGSGRASKVVFVFPGQGSQWIGMGRGL
WESSAVFRDSVLRTDRALAEFVDWSAAAVLAGEPGTPDLDRVDVVQPVLFTVMVALAN
VWRSYGVEPAAVVGHSQGEIAAAHVAGGLSLRDAARVVALRSRALTALAGEGGMVVVQ
RSVAYVENLLRQWEGRLSVAVVNGPEAVVVSGQVAALEELLATEDRARRVAVDYASHS

Query Match 46.2%; Score 246.8; DB 1; Length 32870;
Best Local Similarity 68.0%; Pred. No. 2e-22;
Matches 363; Conservative 0; Mismatches 162; Indels 9; Gaps 1;

Qy 1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
Db 19992 GGAACCATCCTCATCACCGGCGGCACCGGCAACCTCGGCTCCCTCATCGCCCACCACCTC 20051

```
                TDGTVRAARLRRVTASTGASPLSGGTVLVTGGTGGLGRLLVDHLLTAHEAAEVVVASR
                GGRPDGAPADDRVRYVAADVTDRDELAALVDGVAERLCAVVHMAGIVDDAVVATMRPQ
                QWDAVLRVKADVAWQLHELTRDLELAAFVLYSSISATFGGAGQANYATGNAFLDALAR
                HRRDQGLPAVSLAWGLWDEADGMGGRLAATDLTRIARGGMIPMTAEQGLELFDAALHS
                DRAALVPVRLDLAAVAASDQVPPILRDLVPNARRAAAPTAARDMLELVRTSAALVLGH
                RDAHAVEPTRAFKEVGFDSLTGVELRNRLADATGLTLPATLVFDHPTAQALAAHLDEL
                AGARTATRRRTPAPARRHDEPLAIVGMACRLPGGVASPDDLWRLLESGGDGITAFPAD
                RGWDVDALYDPDPEHSGTSTVRHGGFLTGAGDFDAGFFGISPREALAMDPQQRVVLET
                SWEALEHAGIDPHTLRGTETGVFMGGYFYGYGSGADRGDFGATSTQTSVLSGRLSYFY
                CLEGPAVTVDTACSSSLVALHQAGQSLRTGECSLALVGGVTVMASPSGFVDFSQQRGL
                APDGRCKAFAEAADGTAFAEGSGVLVVERLSDAERHGHRVLAVVRGSAVNQDGASNGL
                SAPNGPSQERVIRQALANAGLQPSDVDAIEAHGTGTRLGDPIEATALLATYGQDRATP
                LLLGSLKSNIGHTQAAAGVAGIIKMVLAMRHGTLPPTLHVDTPSSHVDWSAGTVELLT
                DARTWPETGRPHRAAVSSFGVSGTNAHVIIESHPRPTPAPDTGSSTHPVPLLISARTP
                EALDEHTTRVRAFLDAEGEDERAVASALLTRTAFTHRAALIGTDLITGTAEPDRRLVW
                LFSGQGSQRPGMGDELAAAYDVFARTRRDVLDALQVPAGLDVHDTGYAQPAVFALQVA
                LSAQLDAWGVRPDVLVGHSIGELAAAYVAGVWSLDDATELVSARARLMQALPPGGAMA
                AVSASERDALPLLCEGVEIAAVNGPASIVLSGDEDAVLDVAARLGRFTRLRTSHAFHS
                ARMEPMLDEFRDVAERLTYHEPKLPMAAGADCATPEYWVRQVRDTVRFAEQVAAYDGA
                ALLEIGPDRNLARLVDGIPVLHGEDEARSAMTALARLHTGGVAVDWPEVIGAAPTDLP
                HLPTYPFERTRYWLGSRAAGDAAPAGQLPVAHPVLTAAVMVPGTGDLVLTGRVEATDP
                LATRVRHPGAARAVLLDLRSRRATKRAAAPWRRSPWTPRSALPQSGGLALSVTVAAPG
                EDGRRAVAVHARPGTGEWTEHATGILARVARVASAVPETPSPWPPATARPFDVGELAD
                RLARAGHTDGPALPRPRAAWADDDAVHAEVALADEQHADAERYGLHPALLGAALALAG
                EGADLPSAFDDVRVHATGATTVRVAVTATGIHLADETGSPVATVGAVRRRPLTVEGAV
                PGLLRPDLAEIAELPPTTATTGGLDDPVVPDVVILPAHGTGGGPLGETRDLGARVLAA
                LRSFLTDDRYADAVLAVHTGPGLAPAAAAGLVRTAQAEHPRRIVVVDAEPDTAAPLLA
                AAAGLGEPQVVLREGRAYARRLTPAVPSGDAPELDPDGTVLITGGSGTLAGIVARHLV
                GHYGVRRLLMLSRGGTASDVPGAEVTAVACDIADRDELASALAGIDPAHPLTAVVHTA
                AVIDDGVLTALTPERLDTVLRPKADGAWHLHELTKDQDLAALCSTPRPPECSATRDRA
                TTPRRTLLDELAEQRRAGGLPALSVAWGLWEPESGLTAGAGDRMRRDGVTALSTERGL
                ALLDAALRSPEAALVAADPGGLGSSALLRAPHRGPRRRAAGASALADRIAGLSAADAE
                KAALAVVRECAAAVLGHEGPARIEATATFKELGVDSLTAVRLRNAFTEATDVRLPATA
                VFDFPTPQAVAAKLTAALTGRVPARARATKAVGGHDEPLAIVGMACRLPGGVASPDDL
                WRLLESGGDGITAFPADRGWDVDALYDPDPEHSGTSTVRHGGFLTGAGDFDAGFFGIS
                PREALAMDPQQRVVLETSWEALEQAGIVPGTLRGSDTGVFMGAFSDGYGLGTDLGGFG
                ATGTQTSVLSGRLSYFYGLEGPAVTVDTACSSSLVALHQAGQSLRTGECSLALVGGVT
                VMASPGGFVEFSQQRGLAPDGRCKAFAEAADGTAFAEGSGVLVVERLSDAERHGHRVL
                AVVRGSAVNQDGASNGLSAPNGPSQERVIRQALANAGLQPSDVDAIEAHGTGTRLGDP

  Query Match          47.3%;  Score 252.6;  DB 1;  Length 28732;
  Best Local Similarity  67.2%;  Pred. No. 4e-23;
  Matches  357;  Conservative    0;  Mismatches  174;  Indels    0;  Gaps    0;

Qy          4 ACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACC 63
              ||| || || |||| | |||| | || | || | | ||||| || | ||| | ||
Db      13094 ACCCACCACAACACCACCACAACCACCCCCAACACCCCACCCCTCAACCCCAACCACGCC 13153

Qy         64 ACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCAC 123
              | || |    || |      | | |||| || ||     || || |||| | ||||
Db      13154 ATCCTCATCACCGGCGGCTCCGGCACCCTCGCCGGCATCCTCGCCCGCCACCTCAACCAC 13213

Qy        124 GCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGC 183
              | |    | | | |   | | || | |  | |  |    || |||| |  |
Db      13214 CCCCTCACCTACCTCCTCTCCCGCACACCACCACCCCCCACCACACCCCGCACCCACATC 13273

Qy        184 GACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCC 243
              .| |  | | | ||||    ||||     ||    |  || || |  || ||| |||
Db      13274 CCCTGCGACCTCACCGACCCCACCCAAATCACCCAAGCCCTCACCCACATACCACAACCC 13333

Qy        244 CTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACC 303
              |||||| | || |||||||||| | || || ||||||||||||||||||||||||||||
Db      13334 CTCACCGGCATCTTCCACACCGCCGCCACCCACGACGACGCCACCCTCACCAACCTCACC 13393

Qy        304 CCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAA 363
              |||   || |||| || |  |||||   |||||||| ||  ||||   | |||||||||
Db      13394 CCCCAACACCTCACCACCACCCTCCAACCCAAAGCCGACGCCGCCTGGCACCTCCACCAC 13453

Qy        364 CTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTC 423
              | ||||||| ||   |||||||||   |||||||||||||||  |||||||||||| ||
Db      13454 CACACCCAAAACCAACCCCTCACCCATTTCGTCCTCTACTCCAGCGCCGCCGCCACCCTC 13513

Qy        424 GGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCAC 483
```

```
              |||  |||||||||||||||||||||| |||||||||| ||||||||||||||||  |
Db      13514 GGCAGCCCCGGCCAAGCCAACTACGCCGCCGCCAACGCCTTCCTCGACGCCCTCGCCACC 13573

Qy        484 CACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              |||||||||||||    | |||||||||| ||||||||||||||  |||
Db      13574 CACCGCCACACCCAAGGACAACCCGCCACCACCATCGCCTGGGGCATGTGG 13624


RESULT 13
AF007101
LOCUS       AF007101               32870 bp    DNA     linear   BCT 13-JAN-1998
DEFINITION  Streptomyces hygroscopicus putative pteridine-dependent
            dioxygenase, PKS modules 1,2,3 and 4, and putative regulatory
            protein genes, complete cds and putative hydroxylase gene, partial
            cds.
ACCESSION   AF007101
VERSION     AF007101.1  GI:2624946
KEYWORDS    .
SOURCE      Streptomyces hygroscopicus.
  ORGANISM  Streptomyces hygroscopicus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 32870)
  AUTHORS   Ruan,X., Stassi,D., Lax,S.A. and Katz,L.
  TITLE     A second type-I PKS gene cluster isolated from Streptomyces
            hygroscopicus ATCC 29253, a rapamycin-producing strain
  JOURNAL   Gene 203 (1), 1-9 (1997)
  MEDLINE   98085969
   PUBMED   9426000
REFERENCE   2  (bases 1 to 32870)
  AUTHORS   Ruan,X. and Lax,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-1997) 47PAP9A, Abbott Laboratories, 100 Abbott
            Park Road, Abbott Park, IL 60064, USA
FEATURES             Location/Qualifiers
     source          1. .32870
                     /organism="Streptomyces hygroscopicus"
                     /strain="ATCC 29253"
                     /db_xref="taxon:1912"
     CDS             complement(108. .1130)
                     /note="orf5; similar to RapK"
                     /codon_start=1
                     /transl_table=11
                     /product="putative pteridine-dependent dioxygenase"
                     /protein_id="AAC38060.1"
                     /db_xref="GI:2624947"
                     /translation="MNPSSLVLNGLTSYFENGRARVVPPVGRNILGVVNYASVCEYPT
                     LDHGYPELEINMVAPTAEPPAEVWVTDAESEHGERDGITYAHDGEYFFCAGRVPPTGR
                     YTEATRAAYVTMFELLEEFGYSSVFRMWNFIGDINRDNAEGMEVYRDFCRGRAEAFEQ
                     CRLEFDQFPAATGIGSRGGGIAFYLLACRSGGHVHIENPRQVPAYHYPKRYGPRAPRF
                     ARATYLPSRAADGVGGQVFVSGTASVLGHETAHEGDLVKQCRLALENIELVISGGNLA
                     AHGISAGHGLTALRNIKVYVRRSEDVPAVREICREAFSPDADIVYLTVDVCRSDLLVE
                     IEGVVM"
     CDS             1454. .9625
                     /note="orf1; loading domain"
                     /codon_start=1
                     /transl_table=11
                     /product="PKS module 1"
                     /protein_id="AAC38061.1"
                     /db_xref="GI:2624948"
                     /translation="MLRRELIRPLPELLKAHAARRGSQTAFADSRRGVSYAQLEARTR
                     RLAGHLAQLRLLPGDRAAIYLGNCVETIESYLAIARASAIGVPINPHSSTAELAYLLD
                     DSAARVIVTDHTRLPQVRRLQAERPHLTVVVTGDEESGAVPFDALAETEPQQPARDDL
                     GLDDIAWMLYTSGTTGRPKGVLSTQRSCLWSVAASSAGVLGLSETDRVLWPVPLFHSL
                     AHIYCVLSVTAVGATARITEAFDAEELLHTLRAEEFTFLAGVPTMYHYLLDAARDGDL
                     SLPNLRVCLSAGAISTATLRAEFQETFGVPLLDCYGSTETCGLITANWPNGTQAEGSC
                     GLPVPGLTVRIVDPETHESVETGEEGEIWVGGPSLMVGYHNQPEATEAALPRGWYRTG
                     DLGRRDDLGYHTITGRLSELIIRSGENIHPTEVEQVLLRVPGVTDAAVVGSAHPTLGE
                     VPVAFLVPAADGFDPDEVFAACREHLAYFKVPEELYEIGSIPRTGSGKIKRHALAQSP
                     ARLRAVSSGSFESLFRLEWTAPVPEAGPVDGDQATSAHRAGGSVTAGLEAPGSAVVDC
                     PRGSLDEVLRAVRTGLTDRSGRVAVLTHGADHDPAQAAVWGLVESMRADAPHDAPVLV
                     DADGGEATAALSAVLASGQPRALVRGGEVLVPRLARAGSGQRVSRPDPDGTVLVFGAD
                     QEQAAELAGHLVAAYGIRHVRLLARTAVPDLVARLTELGASVTQTVRAPGDRASLAEA
```

```
  JOURNAL   Eur. J. Biochem. 256 (3), 528-534 (1998)
  MEDLINE   98451508
   PUBMED   9780228
REFERENCE   2  (bases 1 to 28732)
  AUTHORS   Motamedi,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-AUG-1998) Molecular Pharmacology/Immunology, Merck
            Research Laboratories, P.O.Box 2000 Bldg. RY80M-213, Rahway, NJ
            07065, USA
FEATURES             Location/Qualifiers
     source          1. .28732
                     /organism="Streptomyces sp. MA6548"
                     /strain="MA6548"
                     /db_xref="taxon:82632"
     gene            complement(61. .4575)
                     /gene="fkbP"
     CDS             complement(61. .4575)
                     /gene="fkbP"
                     /codon_start=1
                     /transl_table=11
                     /product="FK506 peptide synthetase"
                     /protein_id="AAC68816.1"
                     /db_xref="GI:3798625"
                     /translation="MTSRGKRPGTLRLGGSPKKRNSTNREIVAADRAEPLLASFAQRR
                     IWFLQQLDPDSNAYNLPLVRRLTGPLDAALLERALTLVVARHEALRTVFDAPDGEPVQ
                     RVLPAPPRILRHARADGEQGAARLVHDEIAAPFDLSTGPLIRALLIRVGDDDHVLAVT
                     VHHVAADGWSFGLLQLELAAHCTALRDTSRPAVLPPLPLRYADFAAWERRELTGAGLE
                     KRLAYWRDRLRGAPAVLDLPTDRPRPPVADADAGLVEWRPPAELAAAVRRIGRDAGTS
                     AFMTLLSAFQLVLARHAGTRDVLVGTPVANRTRTEFEGVIGMFVNTLALRGDLSGDPS
                     FRELLDRCRATAADAFAHADLPFENLIEVLAPDRDLSVNPVVQVLLQVLRRDASAPAL
                     PGVTAEPFGTGRWFTRFDLEFHVYEEADGALTGELLYSRALFDESRITRLLDEFTAVL
                     RAATAAPGTRLSQLPGAEATPTPEIPSNDTARELPVDTLPELLARYAARTPDAVAVTD
                     PCTSLTYGELDRRANRLAHLLRARGTVTGDLVGICADRGVDLVVGIVGILKAGAAYVP
                     LDPEHPPERTDFVLADARLTAVVAHETYRHRFSGVPHVVTPDDPDLGRQPDTAPHLTL
                     GRDSLAYAIYTSGSTGKPKAVLMPGVSAVNLLLWQERTMGREPASRTVQFVTATFDYS
                     VQEIFSALLGGTLVIPPDEARFDPPGLARWMDEQAITRIYAPTAVLRALVEHVDPHSD
                     QLSALRHLCQGGEALALDARLRELCRHRPHLRVHNHYGPAESQLITGYTLPADPDTWP
                     AAAPIGRPIDNTRIHLLDDALRPVPDGMPGQLCVAGIGLARGYLARPELTAERWVSGG
                     ATGEERMYLTGDLARRAPGGDLDFLGRIDDQVKIRGIRIEPGEVENVLAEDARVAHAA
                     VSVREDPRGEKFLAAYVVPTAGPHGDGLAASLREGLADRLPAYLVPSAVVVMDALPRT
                     TSGKVDRRALPDPAPGPASPRKVAPRNRAERVVCRIFQDVLDVPQVGIDDEFFALGGH
                     SLLATRVVSRIRAELGADVALRTLFDARTPAALARAVDEAGPAALPPILSSPCAGPAP
                     ITAAQEQMLHSHGSLLAAPSYTVAPYGFRLRGPLDRDALDAALTRITARHEPLRTGFR
                     GREQIVGPPAAVCAEVVEVPSGDVGAAARVARAELARPFDLVAGPLLRAVLLPLGAED
                     HVLLLMLHHLAGDGWSFDLLVRELSGASPDLPVSYPDVARWERVPAVVAARENDRAYW
                     RRQFEGACAPELPTVRPRGTAAPDGRAFLWTLDDATVAAARRVADARNATWHETVLGA
                     FALVVAEAADTDDVLVATPFADRGYAGTDHLIGFFAKVLALRIDVGSGTGGTPSFAEV
                     LRRVHATMTGAHAHQSVPYSAMRAADPALPPTPVSFQLISALSTELRLPGLRTEPFPV
                     VTETVDEINGELSVNLFDDGHTVSGAVVHDAALLDRATVDDLLTRVESTLRAAAGDLT
                     VRVTGQVGSE"
     gene            complement(4572. .5570)
                     /gene="fkbO"
     CDS             complement(4572. .5570)
                     /gene="fkbO"
                     /codon_start=1
                     /transl_table=11
                     /product="FK506 oxidase".
                     /protein_id="AAC68817.1"
                     /db_xref="GI:3798626"
                     /translation="MSAPVAAPYCRFEKIGSPDLEGDETVLGVIEHGTGHTGASLVDG
                     VPHTALHTTTRDEEAFAEVWHAQPPVESGTDDGIAWARTDEYLFGVGRVPESGRYADA
                     VSALYTRISGLTRSLGHPLLARTWNYVSGINSANADGLEVYRDFCVGRAQALDEGGID
                     PASMPAATGIGTHGGGIACAFIAARGGVRVNIENPAVLTAHHYPAAYGPRSPVFARAT
                     WLGPPDRGRLFVSATAGILGHETTHHGDVTGQCEVALDNIARVVGAENLRRHGIRRGH
                     LLADVDHLKVYVRRREDLAAVRRVCAARLSSTATVAVLHTDIARGDLLVEIEGVVR"
     gene            5798. .28528
                     /gene="fkbB"
     CDS             5798. .28528
                     /gene="fkbB"
                     /function="FK506 polyketide backbone synthesis"
                     /note="initiation and first four elongation steps"
                     /codon_start=1
                     /transl_table=11
                     /product="FK506 polyketide synthase"
                     /protein_id="AAC68815.1"
                     /db_xref="GI:3798624"
                     /translation="MARAESIRPLHELLRVNAERHGDRVAYTDSRRAVTYGRLQLRTG
                     RIAGHLAASGVGRGDRVRLLLGNRIETIEVYLAAARAAAVAVPLNPDAGDAELAHFLT
                     DSGATVLVTDDLHLDQVRRVGTDATVVLAGRREPGCVAYEDLAGTEPPRPPRDDLGLD
                     EPAWMLYTSGTTGRPKGVVSAQRSGLWSATYCDVPSWELTEDDELLWPAPLFHSLGHH
                     LCLLAVLTVGASARILGGFVARDVLDALAEHPCTVLVGVPTMYRYLLGAVSGEPQTRA
                     LRVALVAGSTSPASLTAGFEAAPGVPLLDTYGCTETTGSLTANTLTDTRVPGSCGLPV
                     PGLSLRFVDPVTGADVRRGDEGELWASGPSLMLGYHAQTEATAQVLADGWYRTGDLAR
                     QAETGHVTITGRVKELIIRGGENIHPREIESVAQEVPGVRDAAAAGRPHPVLGEIPVL
                     YVVPEGPVVPADAILAECRRQLAYFKVPDEILHVTTVPRTASGKVRRGELAGLPARLV
                     GTGSGEAALCELVWERRELPDTVSQVQAVVTRRAVGIGPAEVPDADQALRWDEALVRQ
                     AAEPGSFVLVDTDTDTDTDANNGDEADELVAAAAALGEPQVALRSGVAYVPRLVRADT
                     TPLPASQWALRPPASGTLRDLAVAPSDAPPRPLAAGEVRVDVRAAGLNFRDVLIALGT
                     YPGDGAMGGEAAGIVTEVGPGVDDLVPGAHVFGLVQDAFRRSVVADRRLVARIPRGWS
                     FPVAASVPIVFATAWYGLVDAGQVRPGQKVLIHAATGGVGSAARQIARHMGADVYATA
                     SPAKQHLLLADGFDADHVANSRTTAFADAFPPVDVVLNSLTGEFLDASVELLAPGGRF
                     IEMGKTDIRHDVQQPFDLSDVDPARLREILELLLELFDRDELRPLPVRPWDIRRARDA
                     FGWMSQARNTGKTVLTVPRPLGPDAPVLVTGEGADAVAGLLRDETGRTHVSSSAAEAE
                     RAEEGEEGVNASTLAHVLLGDPAGADAYARSRASEGLPIVTVHGLPGIEQAALTGELI
                     EAAIEGGGSYVVTRVGSAGARAEAMAGATPPILSGLAGPTGRAEPGAAGPEWGSRLAA
                     ADADRENVLLDLVRDSVATVLGLPGAERCAPDRTFRENGLDSLTTVEFANTVAARTGL
                     RVPASTAFDHPTPRSFAAHLAGSVAAPAAKAALGSGEPVAIVGMACRLPGGVTSPEGL
                     WRLVASGTDAITEFPVDRGWDVDELVDPDPDAVGRSTTRHGGFLTEATGFDAAFFGIS
                     PNEALAMDPQQRLVLETSWEAFEHAGIVPDTLRESDTGVFMGAFHQGYGAGRDLGGLG
                     VTATQTSVLSGRLSYFYGLQGPAVTVDTACSSSLVALHQAAQALRSGECSLALAGGVT
                     VMATPGSFVEFSRQRGLSPDGRCKAFADAADGTGFAEGVGVLVVERQSDAERNGHTVL
                     AVVRGSAVNQDGASNGLSAPNGVAQQRVIRQALANAGLSGADVDAVEAHGTGTVLGDP
                     IEAQAVLASYGQEREVPLLLGSVKSNVGHTQAAAGVAGVIKMVMAMRHGLLPKTLHVD
                     EPSSHVDWSAGAVEVVTETRPWPESGGARRAGVSSFGVSGTNAHVILEGVAESSVRDG
                     GSSGLVPLPVSARTESSLALQVERLGEYVRGGADAAAVADGLVRGRAVFDRRAVLLGE
                     SAVAGVAVEGARTVFVFPGQGSQWVGMGRELMGASEVFAARMRECAAVLEPHTGWDLL
                     DVLGEAVVVDRVEVLQPASWAVAVSLAALWQAHGVVPDAVVGHSQGEIAAACVAGALS
                     LEDAARVVALRSQAIAARLAGRGAMASIAVPASAVETVEGVWIAARNGPESTVVAGDP
                     AAVERVLARYEAEGVRVRRIAVDYASHTPHVEAIEAQLADALEGITSSTPSVPWWSTV
                     DSGWVTEPFGDAYWYRNLRQPVAMDTAVSELDGSLFIECSAHPVLLPALDQERTVASL
                     RTDDGGWDRFLAALAQAWTQGADVDWTTLIEPAPHRVLDLPTYPFDHKRYWLQPAPVT
                     GAGIGHPFLSSAAVLPGSDGVLLTGKLSLATHPWLADHTVQEVVLLPGTAFLELVVRA
                     GDEVGCDTIDELVIETPHSLPASGAVDLTVTVDQPDDTGHRPVSVHARPAGTDTWTRH
                     ATGTLGRTEDTAPDTNSFPQWPPTGAQPASLDEFYERLAGAGYTYGPAFQGLRAAWRA
                     GDTVYAEVALDGAADVDRFGVHPALLDAALHAGRLDAGDELELPFSWTGVRLHATGAA
                     AVRVAHSRGPAGVAVQVADLDGRPVVSVDALVLRPRSVAPSGPNLLGLEWLPVAEAHY
                     DGADELPEGYTLITATHPDDPDDDPTNPHNTPTRTHTQTTRVLTALQHHLTTTNHTLI
                     VHTTTDPPGAAVTGLTRTAQNEHPGRIHLIETHHPHTPLPLTQLTTLHQPHLRLTNNT
                     LHTPHLTPITTHHNTTTTTPNTPPLNPNHAILITGGSGTLAGILARHLNHPLTYLLSR
                     TPPPPTTPRTHIPCDLTDPTQITQALTHIPQPLTGIFHTAATHDDATLTNLTPQHLTT
                     TLQPKADAAWHLHHHTQNQPLTHFVLYSSAAATLGSPGQANYAAANAFLDALATHRHT
                     QGQPATTIAWGMWHTTTTLTSQLTDTERQQVRDGFRPLTEAEGNHFLDASLATDAPFV
                     VAAIPAEPARRRVERRTARTGEDSGRDLLAVVCVATAAVLGHADASEIGAATAFKDLG
                     IDSLSGIRLRNSLAEATGVRLSATAVFDHPTPDALAARLAEELRGDDDSQGAAAAAAA
                     PGTVPTVAAAGDEPLAIVAMACRMPGGVDTPEDLWSLVESGGDAITEFPTDRGWDLAA
                     LYDPDPDAIGKVSVRHGGFLSGAADFDAEFFGISPREALAMDPQQRLVLEVSWEAFER
                     AGILPASVRGSDAGVFMGAFTQGYGAGVDLGGFGATGTPTSVLSGRLSYYFGLEGPSV
                     TVDTACSSSLVALHQAARSLRSGECSLALVGGVTVMATTTGFVEFSRQRGLAPDGRAK
                     AFADTADGTSFAEGAGVLIVERLSDATRLGHPVLAVVRGSAVNSDGASNGLSAPNGPA
                     QQRVIERALDDAGLAPGDIDAVEAHGTGTRLGDPIEAQALEAAYGLDRLHPLLIGSLK
                     SNFGHTQAAAGVAGVIKMVLAMRHGVLPRTLHVDEPSRHVDWGGDVRLLRRNEPWPVT
                     GRVRRAGVSSFGISGANAHVVLEAGPPAAPAPLPAADPVPEDVVWPVSARTPDGVRDV
                     AGRLTALTAPAAAIGHSLATTRTAMRHRAVVPARDAEAFARGEEVPGVVRGTADVTDT
                     RAVFVFPGQGSQWDGMGAELLATEPVFARRLGECAEALAPYTGWDLLDVIARRPGAPE
                     LDRVDVVQPASFAMMVALAELWRAHGVAPAAVVGHSQGEVAAACVAGVLTLDDAAKVV
                     ALRSRLVATERAGHGGMVSVPPADFDAAAWAGRLEVAAVNGPASIVVAGAADAVEELL
                     AATPHARRIAVDYASHTAHVESIRGALLDALADLTPGAPEIPFFSTVDEAWLDRPADA
                     AYWYDNVRCPVRFGAAAARLAELGHRVFVEASPHPVLTTALADTLAGHPNTAVTGTLR
                     RGDGGARRFTRSLAELWVRGVPVSWPPGELRGVPLPTYPFRRDRYWVDAEPAGTSGHP
                     LLGSLVERADGEGAFATALFSVRRQPWLADHEVDGRIIVPGSALVELLAEAGARLGTP
                     TIAELTIVAPVVVDEERDTEIQLTVATEVSGRRAVRLHSRTGAGPWTVGATGALSADT
                     RDPEPVEWPPADAEPVDLTGFYDTLPISYGPAFRAMTAMWAGQGRAYASVRLGEQLTD
                     ARYGLHPVLLDAAVHALGTLFADPERRRLAFSWSGVRIHARAATTQRVLLERADGDTI
                     RILATDEHGAPVLDVDGLTVRAAEPGTEALFEVAWVSVPASPVPGWTHLADVPEGDVL
                     EGDAGEGDVGESDPPPVVVLAVEPGDPDSSPGVRARELGRDLLLATVRTWLAAPRWAES
                     RIVVVTRTGDPAQEALGGLVRTAGTEHPGRVGRIEADEITPATVEAGLAAGDESHVRV
```

```
WAVAVSLAALWQAHGVSPDAVIGHSQGEIAAACVAGALSLEDAARIVALRSQTIAAHL
AGRGAMASIALPATAVETVEGVWVAARNGPESTVVAGDPSAVERVLARYEAEGVRVRR
IAVDYASHTPHVEAIQEQLADVLGDITSSAPSVPWWSTVDGGWVTEPAGDDYWYRNLR
QPVAMDTAIGELDGSLFIECSAHPVLLPALDQERTVASLRTDDGGWDRFLTALAQAWT
QGADVDWTTLIAPAPDRLLDLPTYPFDHKRYWIEATGAADLTALGLTDTAHPMLAATA
ALPTNDGTVLTGRISLRTHPWLADHTIQDTVLLPGTAFLELVMRAGDEVGCDTIDELV
IETPLTLPVTGAVDLTVTVEQADATGHRPVSVHARPAGTDTWTRHATGTLGRTEDTAP
DTNSFPQWPPTGPTALDVTEFYAQLAALGYRFVPTFRGLRAAWRAGDTVYAEVALPED
RAADADHFGVHPALLDAALQSGSLLMLESDGEQGVQLPFSWHGVRFHATGATSLRVAI
VPGPDGLRLRAADSENHPVATIDALVTRPPADLAPSDPVLRVGWTAVPVPAEVRPSDA
DVLMLHSDDADPIQETRDLTTRVLDALLTTEGTLIVQVTGGLAAMAAAGLVRTAQSEQ
PGRFFLVETDPGEVLDGARCDMIAALGEPRLRLRGNRFEAARLVRAAAALTPPDTGSW
QLRQSATGSLDDVAIVPTGEPDRPLGAGEVRIAVRAAGLNFRDDTVALGVVADDRPLG
SEAAGVVLETGPGVHDLAPGDRVLGMLAGAFGPVAITDRRLLGRMPDGWTFPQAASVV
TVFATTWYGLVDLAELRPGEKVLIHAAATGVGAAAVQIARHLDADVYATASTPKQHLL
DLEPTHIADSRTTAFADTFPPADVVLNSLTGPLLDASAQLLTPGGRFIEMGKTDIRHH
VQQPFDLMVAGPDRLQQIITQLLDLFANGTLHPLPVHAWDVRQAREAFGWMSSGRHTG
KLVLTVPRPLDPEGAVIITGGSGTLAGHLARHLDHPHTYLLSRTPPPDTTPGTHLPCD
VSDPTQLATALAQIPQPLTGIFHTAATLDDALLHNLTPQRIDTVLQPKADAAWHLHHH
THTHATDLAAFVLYSAVAGVMGSPGQGNYVAANAFLDALAEHRQAMGLPAQSLAWGMW
ADVSALTAKLTDADRGRIRRGGFPPLSAADGMRLFDSAARTRSRSSSRRPSDVTRLDG
PVAPLLRGLVAHRSGPVRAVARNADDEPLAARLAGHTAAEQRRVMQEVVLRQAAAVLA
YGLGEQVAADRPFRDLGFDSLTAVDLRNRLAAETGLRLPTTVVFSHPTAEALATHLLE
LLDAPAALTTGAPLPAVTAAPGTAARDQDEPIAIVAMACRLPGGVTSPEELWRLVESG
TDAITTAPGDRGWDLDALYDPDPDAVGKAYSLRGGFLEGAAEFDAAFFDISPRESLGM
DPQQRLLLETAWEAIERGRINPASLHGREIGVYVGAAAQGYGLGAEDTEGNAITGGST
SLLSGRLAYVLGLEGPAVTVDTACSSSLVALHLACQGLRLGECELALAGGVSVLSSPA
AFVEFSRQRGLAADGRCKSFGSGADGTTWAEGVGVLVLEQLSDAERLGHTVLAVVRGS
AVTSDGASNGLTAPNGLAQQRVIRKALAAAGLTAADVDLVEGHGTGTRLGDPVEADAL
LATYGQNRGEPVWLGSLKSNIGHATAAAGVAGVIKTVQAIGAGTMPRTLHADEPSPAV
DWTAGQVSLLTGNRPWPMTSVPGERPSPRSGSAGRTLTSSLNSTVRSPWRRDRLARNP
GRCLGCCPRGPLPPCGRRPPVCAIIWRPSGPDPLDIGYAQATSRAQFTHRAAEEAASP
DGFRAGLDSVADGAEAPESSAEPPRSARRFLFDGQGAQRVGMGRELHGRFPVFAAAWD
EVSDAFGKHLEHSPTDVFHGEHGDLAHDTLYAQVGLFTLEVALLRLLEHWGVRPDVVV
GHSVGEVTAAYAAGVLTLADATTLIVARGRALRALPPGAMTAVEGSPAEVGAFTDLDI
AAVNGPSAVVLTGAPDDVAAFEREWAAAGRRAKRLDVGHAFHSRHVDGALDDFRGVLE
SLAFGAARLPVVSTTTGRDAAGDLATPEHWLRHARRPVLYADAVRELADLGVNMFVAV
GPSGALASAASENTGGSAGTYHAVLRARTGEESAALTAVAELHAHGAPVDLAAVLAGG
RPVDLPVYPFQHRSYWLAPAVGGGSPTAVPDTGRPTEPDPDDLTVAEIVRRRAAALLG
IADPGDVDADTTFFALGFDSLAVQRLRNQLTAATGLDLPTAVLFDHDTPSALTAYLEE
QLGDGADDDAPTVLALLAEMDSLDAADTAATPAPERAAIADLLDKLSRTWKDHR"
misc_feature    46. .5740
                /gene="fkbA"
                /function="approx. span of module 7"
misc_feature    136. .1345
                /gene="fkbA"
                /function="approx. span of beta-ketoacyl ACP synthase
                domain of module 7"
misc_feature    1609. .2413
                /gene="fkbA"
                /function="approx. span of acyltransferase domain module
                7"
misc_feature    2509. .3028
                /gene="fkbA"
                /function="approx. span of dehydratase domain of module 7"
misc_feature    3868. .4699
                /gene="fkbA"
                /function="approx. span of enoyl reductase domain of
                module 7"
misc_feature    5521. .5716
                /gene="fkbA"
                /function="approx. span of acyl carrier protein domain of
                module 7"
misc_feature    5741. .10497
                /gene="fkbA"
                /function="approx. span of module 8"
misc_feature    5770. .7345
                /gene="fkbA"
                /function="approx. span of beta-ketoacyl ACP synthase
                domain of module 8"
misc_feature    7354. .8176
                /gene="fkbA"
                /function="approx. span of acyltransferase domain of
                module 8"
misc_feature    8275. .8668
                /gene="fkbA"
                /function="approx. span of non-functional dehydratase
                domain of module 8"
misc_feature    9439. .10045
                /gene="fkbA"
                /function="approx. span of ketoreductase domain of module
                8"
misc_feature    10252. .10471
                /gene="fkbA"
                /function="approx. span of acylcarrier protein domain of
                module 8"
misc_feature    10498. .16234
```

Query Match 48.2%; Score 257.4; DB 1; Length 19791;
Best Local Similarity 67.8%; Pred. No. 1.1e-23;
Matches 360; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

```
Qy          4 ACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACC 63
              |||  || || ||||  |  ||||  | ||  | || | | ||||| || | ||| | ||
Db       9400 ACCCACCACAACACCACCACAACCACCCCCAACACCCCACCCCTCAACCCCAACCACGCC 9459

Qy         64 ACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCAC 123
              | || |    ||  |    |  |  |||| || ||      || || ||||  |  ||||
Db       9460 ATCCTCATCACCGGCGGCTCCGGCACCCTCGCCGGCATCCTCGCCCGCCACCTCAACCAC 9519

Qy        124 GCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGC 183
              | ||   |  |  |  |  |  | || |  |  |  |  |  |    || |||| |  |
Db       9520 CCCCACACCTACCTCCTCTCCCGCACACCACCACCCCCCACCACACCCGGCACCCACATC 9579

Qy        184 GACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCC 243
              |  |  |  | | ||||    ||||      ||    |  || || |  || ||| |||
Db       9580 CCCTGCGACCTCACCGACCCCACCCAAATCACCCAAGCCCTCACCCACATACCACAACCC 9639

Qy        244 CTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCCACCCTCACCAACCTCACC 303
              |||||| | || |||||||||| | || ||||||||||||||||||||||||||||||
Db       9640 CTCACCGGCATCTTCCACACCGCCGCCCACCCTCGACGACGCCCACCCTCACCAACCTCACC 9699

Qy        304 CCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAA 363
              |||   || |||| || |  |||||   |||||||| ||  ||||   | |||||||||
Db       9700 CCCCAACACCTCACCACCACCCTCCAACCCAAAGCCGACGCCGCCTGGCACCTCCACCAC 9759

Qy        364 CTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTC 423
              | ||||||| ||   |||||||||  |||||||||||||||||  ||||||||||||| ||
Db       9760 CACACCCAAAACCAACCCCTCACCCACTTCGTCCTCTACTCCAGCGCCGCCGCCACCCTC 9819

Qy        424 GGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCAC 483
              |||  |||||||||||||||||||||| |||||||||  |||||||||||||||||  |
Db       9820 GGCAGCCCCGGCCAAGCCAACTACGCCGCCGCCAACGCCTTCCTCGACGCCCTCGCCACC 9879

Qy        484 CACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              |||||||||||||    |  ||||||||||  |||||||||||||||  |||
Db       9880 CACCGCCACACCCAAGGACAACCCGCCACCACCATCGCCTGGGGCATGTGG 9930
```

```
RESULT 12
AF082100
LOCUS       AF082100               28732 bp    DNA     linear   BCT 27-OCT-1998
DEFINITION  Streptomyces sp. MA6548 FK506 peptide synthetase (fkbP), FK506
            oxidase (fkbO), and FK506 polyketide synthase (fkbB) genes,
            complete cds.
ACCESSION   AF082100 AF082099
VERSION     AF082100.1  GI:3798623
KEYWORDS    .
SOURCE      Streptomyces sp. MA6548.
  ORGANISM  Streptomyces sp. MA6548
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 28732)
  AUTHORS   Motamedi,H. and Shafiee,A.
  TITLE     The biosynthetic gene cluster for the macrolactone ring of the
            immunosuppressant FK506
```

```
Db      20750 GTCGCCACCCACGGCGTCCGCCACCTGCTGCTCGCCAGCCGCCGCGGCCCGGCCGCCGAC 20809

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              |||  |  |||| ||  || |||||          |||  |  | ||  |||| || ||
Db      20810 GGCGCCGACGACCTGACGGCCGAACTCACCGGGCTCGGCGCCACCGTCCACATCGCCGCC 20869

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||  |  | |||| | |   |||    | || |||  |||| ||||  |    |||
Db      20870 TGCGACGTCGCCGACCCGGCCGCCCTCGCCGACCTGCTCGGCACCGTCCCGGCCGGGCAC 20929

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              || ||||||  |||| ||||||||||| ||| || |||||||||  |  ||||  |  ||||
Db      20930 CCGCTCACCGTCGTCGTCCACACCGCCGGCGTCGTCGACGACGGCGTCCTCGGCTCCCTC 20989

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              |||||    |  ||  ||| ||||||  ||  |||| ||| ||  ||||   | ||| |||
Db      20990 ACCCCGCAGCGCCTGGACACCGTCCTGCGGCCCAAGGCCGACGCCGCCTGGCACCTGCAC 21049

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              |    ||||  |||  |  |||   |||||||||||||| ||| |||| ||||||||||
Db      21050 GAGGCGACCCGCCACCTCGACCTGGACGCCTTCGTCCTCTTCTCGTCCGTCGCCGCCACC 21109

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              |||||   ||||| || |||||||||||||| | |||||||| ||| ||||||||||||
Db      21110 CTCGGCAGCCCCGGACAGGCCAACTACGCCGCCGGCAACGCCTTCCTGGACGCCCTCGCC 21169

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              ||  |||  | ||  |  ||| |||||||||  | ||||||||||| | |||
Db      21170 GCCCGGCGCGCCGCCACCGGCCTGCCCGCCACCTCCCTCGCCTGGGGCCCGTGG 21223
```

```
RESULT 11
SPFKBAD
LOCUS       SPFKBAD                19791 bp    DNA     linear   BCT 26-FEB-1997
DEFINITION  Streptomyces sp. MA6548 fkbA gene and partial fkbD gene encoding
            FK506 polyketide synthase and cytochrome P-450-9-deoxo-FK506
            hydroxylase, respectively.
ACCESSION   Y10438
VERSION     Y10438.1  GI:1781343
KEYWORDS    acyl carrier protein; acyltransferase; beta-ketoacyl-ACP-synthase;
            beta-ketoreductase; cytochrome P-450-9-deoxo-FK506 hydroxylase;
            dehydratase; enoyl reductase; FK506 polyketide synthase; fkbA gene;
            fkbD gene.
SOURCE      Streptomyces sp. MA6548.
  ORGANISM  Streptomyces sp. MA6548
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 19791)
  AUTHORS   Motamedi,H., Cai,S.J., Shafiee,A. and Elliston,K.O.
  TITLE     Structural organization of a multifunctional polyketide synthase
            involved in the biosynthesis of the macrolide immunosuppressant
            FK506
  JOURNAL   Eur. J. Biochem. 244 (1), 74-80 (1997)
  MEDLINE   97217427
   PUBMED   9063448
REFERENCE   2  (bases 1 to 19791)
  AUTHORS   Motamedi,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-1997) H. Motamedi, Merck Research Labs, P.O.Box
            2000, Bldg. Ry80Y-225, Rayway, NJ 07065, USA
FEATURES             Location/Qualifiers
     source          1. .19791
                     /organism="Streptomyces sp. MA6548"
                     /strain="MA6548"
                     /db_xref="taxon:82632"
     gene            46. .19308
                     /gene="fkbA"
     CDS             46. .19308
                     /gene="fkbA"
                     /note="multienzyme complex"
                     /codon_start=1
                     /transl_table=11
                     /product="FK506 polyketide synthase"
                     /protein_id="CAA71463.1"
                     /db_xref="GI:1781344"
                     /db_xref="SPTREMBL:P95814"
                     /translation="MPEQDKTVEYLRWATTELQRTRAELAAHSEPLAIVGMACRLPGG
                     VASPDDLWRLLESGGDGIGAFPADRGWETGADGRGGFLTGAAGFDAGFFGVSPREALA
                     MDPQQRVVLETSWEALEHAGIDPRSLNGSDTGVFLGAFFQGYGIGADFDGYGTTSIHT
                     SVLSGRLSYFYGLEGPAVTVDTACSSSLVALHQAGQSLRTGECSLALVGGVTVMASPA
                     GFADFSEQGGLASDGRCKAFAEAADGTAFAEGSGVLVVERLSDAERHGHRVLAVVRGS
                     AVNQDGASNGLSAPNGPSQERVIRQALAHAGLQPSDVDAVEAHGTGTRLGDPIEATAL
                     LATYGQDRATPLLLGSLKSNIGHTQAAAGVAGIIKMVLAMRHGTLPPTLHVDTPSSHV
                     DWTTGDVELLTQARPWPQTNRPHRAGISSFGVSGTNAHVIIESHPRPTPAPDTGSSTH
                     PVPLLISARTPEALDEHTTRIRAFLDDNPGADRVTVAQTLARRTPFEHRAVLLGDTLI
                     TADPNAGSGPVVFVYSGQSTLHPHTGHQLAATYSVFADAWGEVLGHLNADQGPATHFA
                     HQIALTALLRSWGITPHAVIGHSLGEISAACAAGVLSIGDASALLAARSRLMDELPTG
                     GAMVTVLTSEENALRALRPGVEIAAVNGPHSVVLSGDEGPVLDVAQQLGIHHRLPTRH
                     AGHSARMDPLVAPLLEAASGLTYHQPHTAIPEDPTTAAYWARQVRDQVRFQAHAERYP
                     GATFLEIGPNQDLSPVVDGIPTQTGTPEEVQALHTALARLHTRGGVVDWPTVLGSDRA
                     PVALPTYPFQHKDYWLRATAQVDVTGAGQEKVAHPLLGAAVALPGTGGTVLTGRVSLA
                     SHPWLGGHAVNGTVLLPGAAFLELAVRAGDETGSDLLHELVVETPLTLPATGSVAISV
                     EVAEPDDTGRRAVTVHARADGAALWTRHATGFLGKAAATTAETGDPAPWPPAGARQVD
                     VGDVYERFAGIGYAYGSSFQGLLAAWRDGDTVHAEVALPDEQVPDAARFTLHPALLDA
                     AFQAGALSALDAPGEAARLPFSFQDVRIHAAGATRARVTVSRDGDRSTVRMTGEDGQP
                     VAEVGSVVSRPYAGGSGSDDELLRPVWTELPMPAPSAADPRVEVLDADPLDGEPAAAT
                     RELTARVLDSLQHRLSATDDVTLVVRIGTGLAAAATAGLVRSAQAENPGRFVLVEAPP
                     DTPADLLTACAALDEPQLAVRDGVLFAPRLVRMSDPVHGPLPLPDGDWLLAHSASGVL
                     QDVTLAADGTPRRALEAGEVRVDVRAAGLNFRDVLIALGTYDGATALGGEAAGVVVET
                     GPGVADLAPGDRVFGLTRGGIGPTAVTDRRWLARIPDGWTFATAASVPIAFATAWYGL
                     VDLAGLRAGEKVLVHAATGGVAMAATQVARHLGAEVYATAGTGKQHVLRAAGLPGTHV
                     ADSRTTAFRTDFPRMDVVLNSLTGEFIDASLELLDADGRFVEMGRAELRDPATIAPAY
                     LPFDLLDAGTDRIGEILGELLRLFDAGALTPLPVRAWDIRQARDALGVMSRARHIGKN
                     VLTLPRPLDPEGAIVITGGSGTLAGILARHVSAQHVYLLSRSAPPEGTPGIHLPCDVS
                     DRQQLAAALRRIDRPITAVVHTAGALDDGTLPSLTPGRFDTELRPKADGAWHLHELTK
                     DQDLAAFVLYSSAAGVLGNAGQGNYAAANAFLDALAEQRRAAGLPALSIAWGLWEDAS
                     GLTAELSEADRQRMRRSGFRTITARQGMRLYEAARRTGSPVVVAAALDDAPDVPLLRG
                     LRRTTVRRAAVRERSLADRSPCCPTTSAPTPPSRSSWNSTATVLGHLGAEDIPATTTF
                     KELGIDSLTAVQLRNALTTATGVRLNATAVFDFPTPRALAARLGDELAGTRAPVAART
                     AATAAAHDEPLAIVGMACRLPGGVASPQELWRLVASGTDAITEFPADRGWDVDALYDP
                     DPDAIGKTFVRHGGFLDGATGFDAAFFGISPREALAMDPQQRVLLETSWEAFESAGIT
                     PDAARGSDTGVFIGAFSYGYGTGADTNGFGATGSQTSVLSGRLSYFYGLEGPSVTVDT
                     ACSSSLVALHQAGQSLRSGECSLALVGGVTVMASPGGFVEFSRQRGLAPDGRAKAFGA
                     GADGTSFAEGAGALVVERLSDAERHGHTVLALVRGSAANSDGASNGLSAPNGPSQERV
                     IHQALANAKLTPADVDAVEAHGTGTRLGDPIEAQALLATYGQDRATPLLLGSLKSNIG
                     HAQAASGVAGIIKMVQAIRHGELPPTLHADEPSPHVDWTAGAVELLTSARPWPGTGRP
                     RRAAVSSFGVSGTNAHIILEAGPVKTGPVEAGAIEAGPVEVGPVEAGPLPAAPPSAPG
                     EDLPLLVSARSPEALDEQIGRLRAYLDTGPGVDRAAVAQTLARRTHFTHRAVLLGDTV
                     IGAPPADQADELVFVYSGQGTQHPAMGEQLAAAFPVFADAWHDALRRLDDPDPHDPTR
                     SQHTLFAHQAAFTALLRSWDITPHAVIGHSLGEITAAYAAGILSLDDACTLITTRARL
                     MHTLPPPGAMVTVLTSEEEARQALRPGVEIAAVFGPHSVVLSGDEDAVLDVAQRLGIH
                     HRLPAPHAGHSAHMEPVAAELLATTRELRYDRPHTAIPNDPTTAEYWAEQVRNPVLFH
                     AHTQRYPDAVFVEIGPGQDLSPLVDGIALQNGTADEVHALHTALARLFTRGATLDWS
                     ILGGASRHDPDVPSYAFQRRPYWIESAPPATADSGHPVLGTGVAVAGSPGRVFTGPVP
                     AGADRAVFIAELALAAADATDCATVEQLDVTSVPGGSARGRATAQTWVDEPAADGRRR
                     FTVHTRVGDAPWTLHAEGVLRPGRVPQPEAVDTAWPPPGAVPADGLPGAWRRADQVFV
                     EAEVDSPDGFVAHPDLLDAVFSAVGDGSRQPTGWRDLAVHASDATVLRACLTRRDSGV
                     VELAAFDGAGMPVLTAESVTLGEVASAGGSDESDGLLRLEWLPVAEAHYDGADELPEG
                     YTLITATHPDDPDDPTNPHNTPTRTHTQTTRVLTALQHHLITTNHTLIVHTTTDPPGA
                     AVTGLTRTAQNEHPGRIHLIETHHPHTPLPLTQLTTLHQPHLRLTNNTLHTPHLTPIT
                     THHNTTTTTPNTPPLNPNHAILITGGSGTLAGILARHLNHPHTYLLSRTPPPPTTPGT
                     HIPCDLTDPTQITQALTHIPQPLTGIFHTAATLDDATLTNLTPQHLTTTLQPKADAAW
                     HLHHHTQNQPLTHFVLYSSAAATLGSPGQANYAAANAFLDALATHRHTQGQPATTIAW
                     GMWHTTTTLTSQLTDSDRDRIRRGGFLPISDDEGMRLYDAAVGSGEDFVLAAAMDPAQ
                     PMAGDVPPILSGLRKSARRTARTGQTPAQRLAELPAADRDTALVTLVSDATAAVLGHA
                     DASGIAPTTTFKDLGIDSLTAIELRNRLAEATGLRLSATMVFDHPTPRVLAAELRTDL
                     FGTAAPLPARTARTRHDEPLAIVGMACRLPGGVTSPEDLWRLVASGTDAITEFPTDRG
                     WDIDRMFDPDPDAPGKTYVRHGGFLSEAAGFDAAFFGISPREAWAMDPQQRVILETVW
                     EAFENAGIVPDTLRGSDTGVFMGAFSHGYGAGDDLGGFGATATQNSVLSGRLSYFFGM
                     EGPAVTIDTACSSSMVALHQAAQSLRDGECSLALAGGVTVMPTPLGYVEFCRQRGLAP
                     DGRAKAFAEGADGTSFSEGAGVLVVERLSDAERNGHTVLALVRSSAVNQDGASNGISA
                     PNGPSQQRVIRQALDKAGLAPADVDVVEAHGTGTPLGDPIEAQAIIATYGQDRDTPLY
                     LGSVKSNIGHTQTTAGLAGVIKMVMAMRHGLLPKTLHVDEPSSHVDWSAGAVELLTEA
                     RPWPDSDRPRRAGVSSLGISGTNAHVILEGVAESSVRSVESSGLVPLPVSARTESSLA
                     LQVERLGEYVRGARDLAAVADGLVRGRTVFGHRAVLLGESTVAGVAGAESRTVFVFPG
                     QGSQWVGMGRELMGCSEVFAARMRECAAVLEPYTGWDLLDVLGEAVVAERVEVLQPAS
```

```
Db      39661 GTCACCGAGCACGGCGCCCGCCACCTGCTGCTGGCCGGCCGGCGCGGCCCCGACGCGCCC 39720

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180

Db      39721 GGCGCCGCCGCACTCCACGCCGAACTGACCGCCCTGGGCGCCGAGGTCACCGTCGCCGCC 39780

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240

Db      39781 TGCGACGTCGCCGACCGCACCGCGCTCGCCGCGCTGCTCGCCACCGTGCCCGCCGAACAC 39840

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300

Db      39841 CCCCTCACCGCGGTCGTGCACACCGCCGGCGTCCTGGACGACGGCACCCTCACCGCCCTG 39900

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360

Db      39901 AACCCCGACCGCCTCGCCACCGTCCTACGGCCCAAGGTGGACGCCGCCTGGCACCTGCAC 39960

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420

Db      39961 GACCTCACCCGCCACCTCGACCTGGCCGCGTTCGTGCTCTACTCCTCCACCGCCGGCGTC 40020

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480

Db      40021 ATGGGCGGACCGGGCCAGGCCAACTACGCGGCCGGCAACACCTTCCTCGACGCGCTCGCC 40080

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534

Db      40081 GCCCACCGACACGCCCTCGGCCTGCCCGCCACCTCGCTGGCCTGGGGCGCCTGG 40134


RESULT 9
AX211739
LOCUS       AX211739               125401 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 35 from Patent WO0159126.
ACCESSION   AX211739
VERSION     AX211739.1  GI:15523950
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 125401)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 35 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .125401
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT    15664 a  49692 c  42871 g  17174 t
ORIGIN

  Query Match           50.0%;  Score 266.8;  DB 6;  Length 125401;
  Best Local Similarity   68.7%;  Pred. No. 5.1e-25;
  Matches  367;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60

Db      40009 GGCACCGTGCTCCTCACCGGCGGCACCGGCGGCCTGGCCGGCCTGCTCGCCCGCCACCTG 40068

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120


Db      40069 GTCACCGAGCACGGCGCCCGCCACCTGCTGCTGGCCGGCCGGCGCGGCCCCGACGCGCCC 40128

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180

Db      40129 GGCGCCGCCGCACTCCACGCCGAACTGACCGCCCTGGGCGCCGAGGTCACCGTCGCCGCC 40188

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240

Db      40189 TGCGACGTCGCCGACCGCACCGCGCTCGCCGCGCTGCTCGCCACCGTGCCCGCCGAACAC 40248

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300

Db      40249 CCCCTCACCGCGGTCGTGCACACCGCCGGCGTCCTGGACGACGGCACCCTCACCGCCCTG 40308

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360

Db      40309 AACCCCGACCGCCTCGCCACCGTCCTACGGCCCAAGGTGGACGCCGCCTGGCACCTGCAC 40368

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420

Db      40369 GACCTCACCCGCCACCTCGACCTGGCCGCGTTCGTGCTCTACTCCTCCACCGCCGGCGTC 40428

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480

Db      40429 ATGGGCGGACCGGGCCAGGCCAACTACGCGGCCGGCAACACCTTCCTCGACGCGCTCGCC 40488

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534

Db      40489 GCCCACCGACACGCCCTCGGCCTGCCCGCCACCTCGCTGGCCTGGGGCGCCTGG 40542


RESULT 10
AX211705
LOCUS       AX211705               65140 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 1 from Patent WO0159126.
ACCESSION   AX211705
VERSION     AX211705.1  GI:15523937
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 65140)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 1 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .65140
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT     8270 a  25171 c  22273 g   9426 t
ORIGIN

  Query Match           48.8%;  Score 260.4;  DB 6;  Length 65140;
  Best Local Similarity   68.0%;  Pred. No. 3.7e-24;
  Matches  363;  Conservative    0;  Mismatches  171;  Indels    0;  Gaps    0;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60

Db      20690 GGCACCGTCCTGATCACCGGCGGCACCGGCGGCCTGGGCGCGGTCCTCGCCCGCCACCTG 20749

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
```

```
                TLLDAPEQAPDPLRPGTPARAEGRVEFTDVSFRYSPDTPLIENLSLTVEPGSTVAIVG
                PTGAGKTTLGNLLMRFYEPDSGRILLDGTDTATMTRDDLRSRFGLVLQDTWLFGGTIA
                ENIAYGAPGACRADIEEAARATCADRFIRTLPQGYDTVLDDESGTVSAGEKQLLTVAR
                AFLARPAVLVLDEATSSVDTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIVVM
                DAGRIVEQGTHDQLLCAQGLYARLHAARTHTPTAGAAAG"
gene            complement(2662. .4416)
                /gene="nysH"
CDS             complement(2662. .4416)
                /gene="nysH"
                /function="presumably involved in efflux of nystatin"
                /note="putative transporter (ABC family)"
                /codon_start=1
                /transl_table=11
                /product="NysH"
                /protein_id="AAF71764.1"
                /db_xref="GI:8050838"
                /translation="MLLRLLRAQLRPYAWATAALVALQLVQILGTLLLPTLGAALIDQ
                GVVRGDGGRITELGVVMGVVALVQIAAALGAAALAARTATAMGRDLRSALFRRILDFS
                AREIGRFGTPSLLTRSVNDVQQVQNLAQTGFGIVVCAPLMCLGSVLLALRQDVPLALL
                LVALVLVVAVCFGLLLARMGTLYARMQLTLDRLGRLLREAITGVRVVRSFVRDDHERA
                RFAQTNDAFLVVSRRVGRLIATMLPVVLLLMNGFTVALLWTGSHRIDAGRMPIGSLSA
                LLSYLSLILMSVVMLAFVFLSVPRARVCAGRIAEVLDTGSSVAPPAAPQPVRGPAGRI
                ELCAAGYRYPGAEEPVLRDVDLTVEPGERIAVLGSTGSGKTTLLNLVLRLADATEGAV
                RVGGTDVRELTAATLAAAVGFVPQRPYLFSGTVASNLRFGRPDATDEELWEALRVAQA
                ADFVARMPDGLDAEITQGGGNVSGGQRQRLSLARALLRRPEIYLFDDCFSALDQATDA
                ALRTALVPYTAGATVITVAQRISAGRDADRIVVLDRGRVVAQGTHDVLLRTSPTYREI
                ALSQLTEEEAAHGLAGRP"
gene            4714. .5748
                /gene="nysDIII"
CDS             4714. .5748
                /gene="nysDIII"
                /function="presumably involved in mycosamine biosynthesis"
                /note="putative dGDP-mannóse-4,6-dehydratase"
                /codon_start=1
                /transl_table=11
                /product="NysDIII"
                /protein_id="AAF71765.1"
                /db_xref="GI:8050839"
                /translation="MSKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKSRV
                SRLASELDFIDGDLMDQGSLVSAVDTVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                VLRMLEAIRMVSGLSTSRTVSPRGQIRFYQASSSEMFGKAAETPQRETTLFHPRSPYG
                AAKAYGHYITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKQGLQDKL
                ALGNLDAVRDWGYAGDYVRAMHLMLQQDAGDDYVIGTGQMHSVRDAVRIAFEHVGLNW
                EDYVVIDPDLVRPAEVEVLCADSAKAQDRLGWKPDVDFPTLMRMMVDSDLAQVSRENQ
                YGDVLLAANW"
gene            5930. .34363
                /gene="nysI"
CDS             5930. .34363
                /gene="nysI"
                /function="responsible for condensation steps 9 to 14 in
                the nystatin polyketide backbone synthesis"
                /note="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="NysI"
                /protein_id="AAF71766.1"
                /db_xref="GI:8050840"
                /translation="MDNEQKLRDYLKLATADLRRTRRRVHKLESAAQEPVAIIGMTCR
                YPGGVRSPEDLWRMVEAGEHGVTPFPTDRGWDLEALAAAPTASGGFLHDAPDFDADFF
                GISPREAVAMDPQQRVVLESAWEAFERAGIDPTSVKGSRTGVFIGAMAQDYRVGPADG
                AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                AGGVTIMSGPGTFIEMGRQGGLSADGRCRSFGDTADGTGWAEGVGILVLERLSDAVRN
                GHEILAVVRGTAVNQDGASNGLTAPNGPSQQQVIQQALVNARLAAGDIDVVEAHGTGT
                TLGDPVEAQALLATYGQNRPADRPLLLGSVKSNLSHTQAAAGVAGVIKMVMAMRHGTL
                PRTLHAEEPTHHVDWSQGAVRLLTDTTDWPATGAPRRAAVSSFGISGTNAHTIIEQAP
                EPQPEDAATAQDDAAGSTPATAPVVPGVVPVLLSGRTPDALRGQAAALRAALDTGRRP
                DLLDLAHSLATTRAGFEHRAVLLATDHPALTDGLTALADADDPAAAPAWITGTTRAET
                RLAVLFTGQGAQRLGAGRELAARFPAFATALDAALDAFTPHLDRPLREVLWGTDAALL
                DRTAYAQPALFAVEVALYRLIESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAATLVA
                ARGRLMQALPDGGAMIAVQASEADVAPLLAGHEDQVAIAAVNGPSAVVLSGAEATVTA
                LAEQLAADGRKTRRLRVSHAFHSPLMEPMLDAFRAVVEDLTLQPPLLPVVSNLTGKPA
                TVAQLTSADYWVDHVRHAVRFADGIDWLARHDTTAFLELGPDGVLSAMAQDCLDAADA
                DAVTLPALRAGRPEEHTLTTALAGLHVHGATLDWTGCFAGTGARRTDLPTYAFQRRRY
                WPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRG
                TTLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRT
                VNVHARPDTGDDTPWTAHATGVLTTADASRQLPASSEQGGTPLAGDPHPALDAAQWPP
                AGAEPLPLDGHYDRLADGGFGYGPVFQGLRAAWRGGDVVYAEVELPEAGRSDAEAFGL
                HPALLDAALHAAPFTGLGERGRGGLPFSWEGVSLHAGGATTLRVRLTPVADDALALTV
                ADGTGAPVLSVDSLVLRSVATQQLDTAAAVARDALFRLDWTPVQPTATDPGPVALLGA
                DPFGLLTHAGFADAPAYPDLAALAAADGPVPTTVVLSLAGTGDDAADPARSAHRCAAE
                ALAAVQTWLDHHERFAAARLVFVTRGATVGRDVAAAAVWGLVRSAQSENPGCFALVDL
                DPDGAVGAAALVAALVSGEPQLAVRGDVLRVARLVRRPLTEVGAGADGTGDGVGDGSG
                VSFSGEGAVLVTGGTGGLGAVLARHLVAEYGVRDLLLVSRSGERAVGAGELVAELAGV
                GARVRVVACDVTDRAAVVELVGGHAVSAVVHAAGVLDDGMVGALTGERLSAVLRPKVD
                AVWHLHEATRGLDLDAFVVFSSLAGVFGSPGQANYAAANAFLDALMTRRRAEGLPGLS
                LAWGPWEQSGGMTGTLTDVDAERLARSGVPPLSVAQGLALFDAAVAGTDATCVPVRLD
                LPVLRARGEVPPLLRSLIRVRARRAAVAGSATAGNLAQRLRRLDEDGRDEMVLDLVRG
                QVALVLGHATGGDVDAGRAFRDLGFDSLTAVELRNRLNTVTGLRLPATLVFDYPTVRH
                LATYVLDELLGTDAEVATVQPAAVAVADDPIVIVGMACRYPGGVSSPEDLWRVLTEGT
                DAVSGFPTNRGWDVESLYHPDPDHPGTSYTRSGGFLHEAGEFDPGFFGMSPREALATD
                SQQRLLLESSWEAIERAGIDPVSLRGSRTGVFAGVMYSDYSAMLASPEFEGFQGSGSS
                PSLASGRVAYTLGLEGPAVTVDTACSSSLVAMHWAMQALRSGECGLALAGGVTVMSTP
                AVFVDFARQRGLSPDGRCKAFADAADGVGWSEGVGVLVLERQSDAVRNGHEILAVVRG
                SAVNQDGASNGLTAPNGPSQQRVIRQALASGGLTAGDVDVVEAHGTGTTLGDPIEAQA
                LLATYGRDREPERPLLLGSVKSNLGHTQAAAGVAGVIKMVLAMRHGVVPRTLHVDAPS
                SHVDWSEGAVELLSEQAAWPETGRVRRAGVSSFGISGTNVHVIVEQAPGAKAIAAAGA
                ARRTPGAVPVLLSGRGRSALRGQAARLLGHLQARPDAELVDVALSLATTRSRFEQRAA
                VVAQDRDQLIASLGALAADRPDPAVVEGEAAGRGRTAVLFTGQGSQRAAMGRELHEVQ
                PEFAAAFDAVCAVFDPLLDRPLREVVFAEDGSDEAALLDETGWTQPALFAVEVALFRL
                VESWGVRPDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARATLMQALPTGGAMIAIQA
                TEDEIAAHLDDTVAIAAVNGPQSVVISGDEEAAETIAATFAERGRKTKRLRVSHAFHS
                PRMDGMLDAFRIVAEGLTYRAPRIPLVSDLTGRRADDAEVCTAEYWVRHVREAVRFAD
                CVRTLRDAGATTFLELGSDGLLTAMAEDTLGDDHDAELVPMLRAGRAEELAAATALAR
                LQVRGVDVDWAAYLAGTGARRTDLPTYAFQHAYYWPQLPTPAAALAAADPADQQLWAA
                VERGDARELADILGLGEQDLTPLDSLLPALTSWRRGNQEKHLLDTLRYRVEWTRLSKP
                TAPVLDGTWLLVASDATAADQPALLDGLADALGSHGARVRRLLLDDSCADRAVLAERL
                ARTADVDAATQVLSVLPLDERDADDCPPLTRGLALTVALVQALADTGAQGRLWTATRG
                AVSTNPADPVTHPVQAAAWGLGRGVALEHPRLWGGLVDLPQVFDERAGQRLAGILAVK
                DAPDGEDQVALRATGVSGRRLVRHTVEALPTAAEFTATGTVLITGGTGGLGAEVARWL
                ARAGAQHLVLTSRRGPDAPGAAELRAELEGYGPSVSVVACDVADRDALAAVLTALPEE
                LPLTGVVHTAGVGHYGPLDTLSTAEFAGLTAAKLAGAAHLDALLADRELDFFVLFGSI
                AGVWGSGNQSAYGAANAYLDALALHRRARGLAATSVAWGPWAEAGMAADDAVSETLRR
                QGLGLLDPAPAMTELRRAVVRQDVTVTVADVDWQRYAPLFTSARPSALIAGLPEVRAL
                AADERTEQDATGASEVVTRVRALAEPEQLRLLTDLVRTESATVLGHSSADAVPEGRAF
                RDVGFDSLTAVELRKRLGAATGLSLPSTMVFDYPTPLELAQYLRAEILGAVLEVAGPV
                ATGGADDEPIAIIGMACRFPGGVSSPEQLWDLVASGTDAISEFPVNRGWQTGHLFDPD
                PDRPGTTYSTQGGFLHEADEFDPTFFGISPREALVMDPQQRLLLETTWESFERAGIRP
                ETLRSTLTGTFVGSSYQEYGLGAGDGTEGHMVTGSSPSVLSGRLSYVFGLEGPAVTVD
                TACSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNPFIAFSRQRALAKDGRCKAFS
                DDADGMTLAEGVGVVLVERLSDAQRNGHPVLAVLRGSAINQDGASNGLTAPNGPSQQR
                VIRQALANARLAPGDIDALEAHGTGTPLGDPIEAQALFATYGRDRDPESALLLGSVKS
                NIGHTQSAAGIASVIKMVMALRHSELPPTLHADAPSSHVDWSAGTVRLLTQARAWPET
                GRPRRAAVSSFGISGTNAHVLLEQAPVADTPAEERPAVAPVPIAAGVVPWVVTARSAA
                ALRGQAERLLAHAETVGTALPAAGPLDIGLSLVSARARFEHRAVVVPPAGTDPLAALR
                AVATDGPSPVVARGVADVEGRTVFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAAL
                AEFTDWSLVDVLRGVVGAPSLERVDVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQG
                EIAAAVVSGALSLRDGARVVALRSQAIGRALAGRGGMMSVALSVDVLEPRLVEFEGRV
                SVAAVNGPRSVVVAGEPEALDALHARLTADDIRARRIAVDYASHSHQVEDLHEELLEV
                LAELAPRTSEVPFFSTVTGDWLDTARMDAGYWFRNLRGRVRFADAVADLLAAEYRAFV
                EVSSHPVLSMAVQEAIDEAGVPAVAAGTLRRDQGGTDRFLLSAAEVFVRGVDVDWAGL
                FEGTGASRIDLPTYAFQHEHLWAVPPAPEAVAAADPDDAAFWTAVEDGDVSALTAALG
                TDEDSVAAVLPALTSWRRARRDRSTVDAWRYRVAWKPLGGTLPHPSLTGTWLLVTADG
                IDDTDVAGALETYGAEVRRLVLDEECVDRAVLRERLAGAEDVTGIVSVLAAAERTDAV
                PGTSLVLGTALTVALIQALGDAEIDAPVWALTRGAVSTGRADELTAPVQAQVTGIGWT
                AALEHPQRWGGTLDLPAALDARAAQRLAAVLSGALGSDDQLAIRPSGVFTRRIVRAEA
```

```
Query Match             50.0%;  Score 266.8;  DB 1;  Length 123580;
Best Local Similarity   68.7%;  Pred. No. 5.2e-25;
Matches  367;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||   ||| |||||||||| ||||||| ||| |||  ||  ||| ||| ||||||
Db      39601 GGCACCGTGCTCCTCACCGGCGGCACCGGCGGCCTGGCCGGCCTGCTCGCCCGCCACCTG 39660

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
                |||| | ||   | |||  ||||| || ||  || ||||   ||||||| || | |||
```

```
                     WRLGDEVYADVRLPDGSRTGIGGFGLHPALLETALAASCLAGYGSSPEGAVWISTDWK
                     DVRLHATGADRLRVRLAPDADGALNVQLADRAGRPVASVSSVACRAIAVGEVSNALAR
                     NRDSLFHVDWAPFPVAPSAGTERWVSLGADEPGARRYPDVASLARAVAAGESVDVALV
                     GTETGDTGDRVADVHDVTREALALVQDWLAEERLAGTRLVVRTRGGVTTAGDEDVTDL
                     GAAAVWGLLRSAQSENPGRIVLVDVDDEEASARILPAVVASGEPQSAVRDGQVFVPRL
                     GRVPEAPAGAATGAWDRRGTVLITGGTGSLGGLFARHLVVEHGVRNLLLTSRRGRDAE
                     GAAELEAELTALGARVEIAACDAADRAALAAVLADIPAEYPLTGVLHTAGVLDDGLIG
                     SQTPQRLSAVLRPKVDATWNLHELTRDMDLSAFVLFSSIAAVVGGPGQSTYSAANHFL
                     DALAEHRRSRGLAATSLAWGLWSQTSGMTGHLDEADLKRIARSGFRPVESEQGPALLD
                     IALTVGRAGLVATPMDLTALREQVQVPVLLSKLARTPVRRTAQNAEVSSESLAERLAG
                     LGDAEQREAVLGIVLGQAAVVLGHSDATGIDRQRSFTALGFDSLTSVELRNQLGTSTG
                     LRLPPTLVFDHPTPVALAEYLHGEVRAELSADGSASRPGVDFAAEVRLADDVRPADEV
                     ITTAADPKEILLTGASGFLGAFLLRDLMRTTTARVHCLVRGADEAAAMERLKANAEWY
                     RVWDDIDPERVSVVLGDLAEPRLGLDEESFDALARTVDVVYHNGARVHWLLPYETLKA
                     SNVTGTEEVLRLAARHRTVPVHYVSTVGVFDGVREPGVPLRVTDPTGPAESLPSGYLR
                     SKWVAEQVIEVARDRGLPVSVYRVDVISGDRVNGACQTRDFVWLTLKGLIQARSVPKG
                     TEGRFHLLPVDYVSAAITGISRQPGTVGRTFHLFNQSSLALSQCVELLRSLGYELDEV
                     DWDTWTQVVTSGDNALLPLLDAFEMMTSDTDGFYPPIDTAETVAALEGTGIGIPVLTR
                     ELFEKYVAFFVEEGHFPAAG"
     misc_feature    complement(4732. .4935)
                     /gene="SC2C4.03c"
                     /note="Pfam match to entry PF00550 pp-binding,
                     Phosphopantetheine attachment site, score 68.50, E-value
                     2.5e-17"
     misc_feature    complement(4810. .4857)
                     /gene="SC2C4.03c"
                     /note="PS00012 Phosphopantetheine attachment site"
     misc_feature    complement(5248. .5268)
                     /gene="SC2C4.03c"
                     /note="PS00340 Growth factor and cytokines receptors
                     family signature 2"
     misc_feature    complement(7342. .8286)
                     /gene="SC2C4.03c"
                     /note="Pfam match to entry PF00698 Acyl_transf, Acyl
                     transferase domain, score 248.50, E-value 9.5e-71"
     misc_feature    complement(8596. .9867)
                     /gene="SC2C4.03c"

  Query Match             50.0%;  Score 266.8;  DB 1;  Length 24225;
  Best Local Similarity   68.7%;  Pred. No. 7.1e-25;
  Matches  367;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||  |||||||||||||||  |||||||||||||| ||||  |||| |||  |||||||
Db      16508 GGCACCGTCCTCATCACCGGCGCCACCGGCGCCCTCGGCACCGTCCTCGCCCGCCACCTC 16449

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
                ||||||  ||   |||||  |||||  ||||||   |||||||  ||||  ||  ||  |||||
Db      16448 GTCACCCAGCACGGCACCCGCCACCTTCTCCTCCTCAGCCGACGCGGCACCAACGCCCCC 16389

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
                |||||||  |  |||      |||||          |||   |  | ||||||||||| ||
Db      16388 GGCGCACAAGAACTCCAGCATGAACTCACCGCCCTCGGCGCCACCGTCACCATCACCGCC 16329

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||  ||   |  |||    ||   |||         |  |||  ||  |||||||  |  |||||
Db      16328 TGCGACGCCGCCGACCGCCACGCCCTCACCTCCGTACTCGACGACATCCCCGCCGAACAC 16269

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              |||||||||  |||||  ||||||||||  |||  ||||||||||||      |||  ||   |||
Db      16268 CCCCTCACCGCCGTCGTCCACACCGCCGGCGTCCTCGACGACGGTCTGCTCTCCTCGCTC 16209

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
                |||       |  |||  |  |  ||  ||  ||   |  ||  |  |  ||     |||  |||
Db      16208 ACCGAGGAGCGGCTCGCCGATGTGCTGCGTCCGAAGGTCGACGCGGCGTGGAACCTGCAC 16149

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              ||||  |||  |   ||   |  |||     ||  ||||||||||||||||||  |||||  |   |
Db      16148 GAACTGACCAAGGACCACGACCTGGAGGCATTCGTCCTCTACTCCTCCGTCGCCGGCCTC 16089

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              |||||    |||||||  ||||||||||||||  |   |||  |||  |||  ||||||||||||
Db      16088 ATCGGCAACGCCGGCCAGGCCAACTACGCCGCCGGAAACACCTTCCTGGACGCCCTCGCC 16029

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              || ||||||    | | |   || |||||| ||  | ||||||||||||   |||
Db      16028 CAGCACCGCAGGGCACTCGGACTGCCCGCCGCCTCCCTCGCCTGGGGCCTGTGG 15975


RESULT 8
AF263912
LOCUS       AF263912               123580 bp    DNA     linear   BCT 24-MAY-2000
DEFINITION  Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF263912
VERSION     AF263912.1  GI:8050835
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., StrLm,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Biosynthesis of the polyene antifungal antibiotic nystatin in
            Streptomyces noursei ATCC 11455: analysis of the gene cluster and
            deduction of the biosynthetic pathway
  JOURNAL   Chem. Biol. 7 (6), 395-403 (2000)
  MEDLINE   20334850
   PUBMED   10873841
REFERENCE   2  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
            N-7489, Norway
FEATURES             Location/Qualifiers
     source          1. .123580
                     /organism="Streptomyces noursei"
                     /strain="ATCC 11455"
                     /db_xref="ATCC:11455"
                     /db_xref="taxon:1971"
     gene            complement(46. .783)
                     /gene="nysF"
     CDS             complement(46. .783)
                     /gene="nysF"
                     /function="presumably post-translationally modifies ACP
                     domains on PKS"
                     /note="putative 4'-phosphopantheteine transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysF"
                     /protein_id="AAF71762.1"
                     /db_xref="GI:8050836"
                     /translation="MIELILPATVATEAAYDDRPRPGDRLLSSEREVIARAVESRQRE
                     FTTVRHLARRALRRLGHPDRAILPNRRGAPQWPPGIVGSMTHCAGYRAAAVSPAELSA
                     AVSIDAEPNGPLPAGVLNAIALPSERPHLVALAAHRPDVHWDRLLFSAKESVFKAWYP
                     LTQRELDFSEAEIVIDPTQGAFTARLLVPGPLLGGRRVTVFPGRWHSTPALLTTAVHL
                     PAPTPRRDREHRTHLTVNSPLPRPTFG"
     gene            complement(867. .2684)
                     /gene="nysG"
     CDS             complement(867. .2684)
                     /gene="nysG"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysG"
                     /protein_id="AAF71763.1"
                     /db_xref="GI:8050837"
                     /translation="MASPDDLEEERTAPRPARRLVGLLRPHRRSVALAVSMGVGGIVL
                     NAFGPLLLGRVTDLIADGVLGGVPGPAPGIDFAAIGRLLLVLLALYVVASLFMLAQGR
                     LVASAVWRTIHELRRDAREKLTRLPLRHFDRQPAGELLSRTTNDIDNLQQTLQQTLAE
                     LITSIFSLLTMLVLMLVISPSLAVVMLLSVPVSALIAARISKRAQPHYAAQWSANGTL
                     NAHVEEVCTGHALIKGFDRRAAAEERFDACNDAVYRAAAKAQFASGAMEPVMMFVANL
                     GYVLVAVIGAWKVINGTLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGIASAQRVF
```

```
REFERENCE   2  (bases 1 to 24225)
  AUTHORS   Seeger,K.J. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 24225)
  AUTHORS   Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2001) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     On May 9, 2002 this sequence version replaced gi:12274798.
            Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC and Beowulf Genomics
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
            numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon. Gene
            prediction is based on positional base preference in codons using a
            specially developed Hidden Markov Model (Krogh et al., Nucleic
            Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
            of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl. CAUTION:  We may not have predicted the
            correct initiation codon.  Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon).  If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions. Cosmid 2C4.
FEATURES             Location/Qualifiers
     source          1. .24225
                     /organism="Streptomyces coelicolor A3(2)"
                     /strain="A3(2)"
                     /db_xref="taxon:100226"
                     /clone="cosmid 2C4"
     misc_feature    complement(2. .1477)
                     /note="previously sequenced DNA fragment. EMBL:AF113603
                     Streptomyces coelicolor putative acyl-CoA carboxylase
                     complex A subunit (accA1) gene, complete cds."
     misc_feature    2. .119
                     /note="nominal overlap with Streptomyces coelicolor cosmid
                     SCAH10"
     misc_feature    complement(374. .1471)
                     /note="Pfam match to entry PF00289 CPSase_L_chain,
                     Carbamoyl-phosphate synthase (CPSase), score 585.30,
                     E-value 3.8e-172"
     misc_feature    complement(602. .625)
                     /note="PS00867 Carbamoyl-phosphate synthase subdomain
                     signature 2"
     RBS             complement(1485. .1489)
     RBS             1774. .1778
     gene            1786. .3438
                     /gene="SC2C4.02"
                     /note="SCO6272"
     CDS             1786. .3438
                     /gene="SC2C4.02"
                     /note="SC2C4.02, possible secreted FAD-binding protein,
                     len: 550 aa; similar to TR:Q9ZAR8 (EMBL:U80222)
                     Streptomyces peucetius ceasius DnrW, 485 aa; fasta scores:
                     opt: 1323 z-score: 1315.4 E(): 0; 45.3% identity in 470 aa
                     overlap. Contains Pfam match to entry PF01565
                     FAD_binding_4, FAD binding domain. Also contains possible
                     N-terminal region signal peptide sequence"
                     /codon_start=1
                     /transl_table=11
                     /product="putative secreted FAD-binding protein"
                     /protein_id="CAC22143.1"
                     /db_xref="GI:12274800"
                     /db_xref="SPTREMBL:Q9EX55"
                     /translation="MTEVSRRKLMKGAAVSGGALALPALGAPPATAAPAAGPEDLPGP
                     AAAAAGRRPAPAPVHVTRSDPRYQGLISGSNQRWVGNPDYIRVVSSAGQVVRAVQEAV
                     DKGLKIAVRSGGHCDEDFVANRDVRVVIDMAGMDSVTYDRERRAFAVGPGARLGTVYR
                     TLYKRWGVVLPGGTCPTVGAGGHITGGGYGALSRSRGLTVDHLYAVEVVHVDARGRAR
                     KVVATREEDDPNRELWWAHTGAGGGNFGVITRYWLRSPDAVGDDPSTLLPAAPSEVLL
                     SDVSWSWDDLDEASFTRLLRNFTEWHARNSAPDSPGRFLFSQLKTMHKAAGYFRMSTQ
                     VDAAAPDADRLLDDYLAAISEGTGVTYHVGDRYRAPWLYAVTEWSGFVEASVPRWKSK
                     SAYVREVMPEEQLRAVYRQLTRDDYPGPYGMIAIVGFGGKINEVAPGDTATAQRDSIA
                     KMLYCSLWSDPADDALHQRWIREAYEDVYASTGGVPRPGGVNDGCYINYADADLADPA
                     LNRSGIPWHELYFKGNYPRLQRVKATWDPRNVFSHRLGIRKP"
misc_feature         1948. .2535
                     /gene="SC2C4.02"
                     /note="Pfam match to entry PF01565 FAD_binding_4, FAD
                     binding domain, score 6.50, E-value 1.1e-06"
gene                 complement(3514. .9972)
                     /gene="SC2C4.03c"
                     /note="SCO6273"
CDS                  complement(3514. .9972)
                     /gene="SC2C4.03c"
                     /note="SC2C4.03c, probable type I polyketide synthase,
                     len: 2152 aa; N-terminal domain almost identical to
                     TR:Q9L8J0 (EMBL:AF202898) Streptomyces coelicolor A3(2)
                     type I polyketide synthase (fragment), 720 aa; fasta
                     scores: opt: 4356 z-score: 4494.6 E(): 0; 93.8% identity
                     in 721 aa overlap and C-terminal domain similar to
                     SW:HETM_ANASP (EMBL:L22883) Anabaena sp. polyketide
                     synthase HetM, 506 aa; blastp scores: Score= 761 (267.9
                     bits), Expect= 2.2e-74, Sum P(2)= 2.2e-74; Identities=
                     184/491 (37%), Positives= 278/491 (56%). Also highly
                     similar, in this same cosmid, to the N-terminal region of
                     SC2C4.04c and to SC2C4.05c. Contains Pfam matches to
                     entries PF00109 ketoacyl-synt, Beta-ketoacyl synthase,
                     PF00698 Acyl_transf, Acyl transferase domain and PF00550
                     pp-binding, Phosphopantetheine attachment site and matches
                     to Prosite entries PS00606 Beta-ketoacyl synthases active
                     site, PS00340 Growth factor and cytokines receptors family
                     signature 2 and PS00012 Phosphopantetheine attachment
                     site"
                     /codon_start=1
                     /transl_table=11
                     /product="putative type I polyketide synthase"
                     /protein_id="CAC22144.1"
                     /db_xref="GI:12274801"
                     /db_xref="SPTREMBL:Q9EX54"
                     /translation="MSNEQKMVEYLKWVTTDLQKARQRLAELEAGNDEPVAIVGMACR
                     FPGGVASPDDLWRLVMDEADGITDWPTDRGWDIDAIYDPERGKPGKSYTRKGGFLDGA
                     AQFDPGFFGISPREALAMDPQQRVLLETAWETFEQAGIDPATLRGSQVGVFAGVVEQS
                     YLGLEGPEELEGYLMTSKLSSVASGRIAYTFGFEGPAVSVDTACSSSLVALHLAVQSV
                     RKGESTMALAGGSTVSGNPSGFVDFSRQNGLAEDGRIKSFAAAADGTAWSEGVGLLLV
                     EKLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQERVIHQALAGARLNASDVD
                     VVEAHGTGTRLGDPIEAQALLASYGRQRPADRPLYLGSLKSNIGHTVAAAGAGGVIKM
                     IMAMRHGILPKTLNVDEPTPMVDWEAGAVELLTEARPWPETGAPRRAGVSAFGVSGTN
                     AHVILEEPPAEEPEEPADTAVTALPVLPWALSARTAPALRDQARRLLSHVEERPAEDP
                     LDVAYSLAAERSALEHRAVVVGSDREELLTGLRALAEGRTTPTGAVQDTRQTGKTAFL
                     FTGQGAQRTGMGMDLYRTFPAYAHAFDTITAQLDPHLDQPLHHTITTGHHLHHTGNTQ
                     PALFATEVALYRLLETWGITPDYLAGHSIGELTAAHISGILTLQDACTLVTARARLMQ
                     NLPTHGTMIALQATEQEILPHLTGHEHHLTIAAINSPTSLVISGNQTAAQHIATQLQA
                     QGRKTKTLTVSHAFHSPHMDGMLHDFHHTATQLTYHEPTIPIVSTLTGNLATHNDLRT
                     PTYWTDQLRNTVRYTQAIHTLHTAGVTTYTEIGPDATLTPLTQQTLDTTTAIPLLRHN
                     QPETHTLLTAISRLHNRGIPINWKNLFTTTPAKHTPLPTYAFQHERFWVETSGVAADA
                     SDLGLESAGHPMLGAAVTVAGADQVLFTSRLSLRTHPWLADHVVLDQTLVPATALVEM
                     AIRAGDELGSTVLNDLSVLGPLVLPGKGGVHVQVSVGSPDASGLRELSIYSRPDNADV
                     PWTLNARGRLSFRGPDAPFSLAQWPPAGAEEVPLDEAYDKLAVQGYVYGPVYRGLEHV
```

```
                SSTPGALPAAPMYHIDWQPVDSTSLVDTRRYAVVGADADETAEVLRAFLPDAHTTAHA
                DLAALRAAPDAVAPTPGPALVVALIPGSPPPAPEATSAPDPVRATLAHGLSLVQALIE
                DERLTTGTTLAVVTRGAVEATQTDVPDLAGAALWGLVRSAQSEYPGRFRLIDVDDSPE
                SRAALATALASDEPQLALRAGVMLAPALTPAEAAGSPAPADAAPTAHGMDGTVLLTGC
                TGALGRRVAVHLARRHGVRRMLLVSRSGPNAPESAALERELTELGVHTSFVSCDLADP
                ADVRKAVAGVPSDHPLTGVVHTAGVLDDGALARLTPERIDTVLRAKADAVRHLHEATL
                GHPLRAFVLFSAAAGLLGRPGQASYAAANAVLDAFARLRRAEGLPAVSLAWGLWDERE
                GMAGGLDATALRRLRRDGILPMPADQALELLDQALTTYRDGPALLVPLLLDSTALRRT
                AGENGPASVPPLLRSLVPSHPRRHRTTTATSVRDESGPGPSARLAALPAKERMTVLLD
                TVSEQVAGVLGHASAAEIDPERPFHEIGFDSLATLELRNRLGRLVGLRLPSTLAFDHP
                TPRKVAEWINAELPAANPALGTSAEDPVLTGIDQMARAIALMDAEDGRRADVRQRLVG
                LLTALDRPASGADSTRAAATTVADRLDEATDDEIFAFLDEQLEEQPDEPLGAPLDERL
                DEQP"
gene            13693. .19212
                /gene="nidA2"
CDS             13693. .19212
                /gene="nidA2"
                /note="orf2"
                /codon_start=1
                /transl_table=11
                /product="polyketide synthase module 3"
                /protein_id="AAC46025.1"
                /db_xref="GI:2558839"
                /translation="MTAENDKIRSYLKRATAELHRTKERLAELESARHEPVAVVGMAC
                RYPGGVATPDDLWELVASGTDAISTFPTDRGWDLDGLYDPDPSTPGKSYVRHGGFLHD
                AAQFDAEFFGISPREATAMDPQQRLLLETSWEALEHAGIAPNTLRTTRTGVFTGVMYD
                DYGSRCFPSLPEVEGYVVNGSAGSVASGRVAYTLGLQGPALTVDTACSSSLVALHLAV
                QSLRRGECDLALAGGSTVMSTPTVFVEFSRQRGLAPDGRCKSFADTADGTAWAEGVGV
                LLVERLSDAERLGHRVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALSDAGLSAD
                DIDAVEAHGTGTPLGDPIEAGALLATYGHPERQTPVWLGSLKSNIGHTQAAAGIAGII
                KMIQAMHHGTLPRTLHVDHPSSKVDWKAGSLQLLAQARPWPADPDRPRRAAVSAFGVS
                GTNAHAIIEEPSTLIEAPVPDETPVPAHTPPVIAWPLSAHTPTALRAQAARLDSWLHH
                TQADPQDIAHALATTRTHFKHRAVITGRTRTELHTKLHTLDAIQGTAHPHPRLTLLFT
                GQGAQHPGMGQELYTTDPHFAAALDEVCEELQRCGTQNLREVMFTPDQPDLLDRTEYT
                QPALFALQTALYRTLTARGTQAHLVLGHSVGEITAAHIAGVLDLPDAARLITARAHLM
                GQLPHGGAMLSVQAAEHDLDQLAHTHGVEIAAVNGPTHCVLSGPRTALEETAQHLREQ
                NVRHTWLKVSHAFHSALMDPMLGAFRDTLNTLNYQPPTIPLISNLTGQIADPNHLCTP
                DYWIDHARHTVRFADAVQTAHHQGTTTYLEIGPHPTLTTLLHHTLDNPTTIPTLHREH
                PEPETLTTALATLHTTGHTTTLHTTSPQSHHLDLPTYPFQRDRYWMAVPPRAAVGDLA
                AAGVTAVGHPLLAAVTEVPETDTTVWTGMLSARALPWLADHLVWDRGVAPGTAVLEIV
                LEIGERIGLPCIGELVLEAPLSWGADAPAVVRTVVGALSPEGTREVSLYSRPEDEAVW
                TRHASGRLARAGSEEMGSGVFAELSGVWPPPGAQEVDISGQYELFAAAGVGYGGAFRG
                LRAVWRREDEIFAEVRLPDDHATDAAGYGLHPALLDAALHPIALLDPLGDGGHGLLPF
                VWTDVLRPGNGGHTLRVRVTRAQDPADGGQDDRPGGAVSLAAADRDGVPVFVARSLVL
                RRVSATQLPAAPFVPLHRLDWVPIPGFTESTGAGGSDPSVLDLTELTADVHLTVREVL
                EAVQAHLREPVGDGGRLVVLTRGAAGPEGGADAPAGAAAWGLVRSAQSEHPGRFVLID
                VDGTEESSRALPAVMAAVGDVDAPQVAVRQGRPSVPRLVPLTADRTEAEPGGPGSAGL
                LGPDGTVLITGGTGALAAVTARHLVARHGARHLLLVSRRGTDAPQAAELASELGALGA
                DVEFRACDVTDRAAVQRLLAGIPAEHPLSCVVHTAGVLDDGVISAQTPERVAAVLGPK
                ADGALHLDELTRDRERSVPLVLFSSVSGVLGSTGHAGYAAANAFLDALAARRRADGHP
                ALSLGWGWWTGTGMAAGLDRADTARIRRSGIAPLDAETAMDLFDRALARPEPALLPVR
                LDHAALREAARAGTLPEVLRDLAGVPSRAHPAPGSEGTATSTGTGSGLAERLARRTVA
                ERLPLLLELVRTEVAAVLGHGDPGAVGAERSFKDAGFDSLTAVDLRNRLNARTGLRLP
                ATLVFDHPTPLALAELLLAGLDVEPAADGPEDLAAAFDRLARGLAALDDRAARVGAAE
                RLRSLLTALTSADPATGTGAQAGPEADGALTEVVTDRLRSASDDELFDLYDSEFR"
gene            19269. .30233
                /gene="nidA3"
CDS             19269. .30233
                /gene="nidA3"
                /note="orf3"
                /codon_start=1
                /transl_table=11
                /product="polyketide synthase modules 4 and 5"
                /protein_id="AAC46026.1"
                /db_xref="GI:2558840"
                /translation="MSATNEEKLREYLRRAMGDLHSARERLAELESARHEPVAVVGMA
                CRYPGGVATPDDLWELVASGTDAISTFPTDRGWDLDGLYDPDPSTPGKSYVRHGGFLH
                DAAQFDAEFFGISPREATAMDPQQRLLLETSWEALEHAGIAPHTLRTTRTGVFTGVMH
                HDYGSHQVGSAADASGQLGLGTAGSVASGRVAYTLGLQGPALTVDTACSSSLVALHLA
                VQSLRRGECDLALAGGTTVMATPTVFVEFSRQRGLAPDGRCKPFADTADGTAWAEGVG
                MLLVERLSDAERLGHRVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALADAHLTP
                DDIDAVEAHGTGTPLGDPIEAGALLATYGHPKRQTPVWLGSLKSNIGHTQAAAGIAGI
                IKMVQALRHDTLPRTLHADHPSSKVDWEAGPLQLLAQARPWPADPDRPRRAGISAFGV
                SGTNAHAIIEEPPTLTEAPADTHTPPVIAWPLSAHTPTALRAQAARLRARLHRTSDAL
                SLADSAGIGHGLAVGRAALPHRAVLLGDGAAPLDALAALASGEVSPDVVTGSAADVRR
                VAFVFPGQGAQWAGMGAELLDSSPVFAAELARCEAALEPFVDWSLTDVLRGAPGAPGL
                DRVDVVQPVTFAVVVALAAMWRWLGVEPAAVVGHSQGEIAAAHVAGVLSLEDAARVVA
                LRSQLIARELAGRGSMASVALAAADVESRLAGAEAGGGVRDVEIAAVNGPETTVVCGA
                PGAVDSLLGVLQGEGVRVRRIDVDYASHSRHVEGIRDELAAVLAGLRPRAGRVPFYST
                VEAEPLDGTALDAGYWYRNLRQRVRFESALRAMLADGVDAFVECSPHPVLTVPVRQTL
                EDAGAGAVAVGSLRRDDGGLRRFLTSAAEAQVAGVPVDWAALCPRAGWVDLPTYAFQR
                ERYWVAPAEPGPAAGAGSAAATGPAAATAARSGSGQEDAASAYGDRLAYHVAWQGLLP
                GADAWRPGPRLLIVPAGEREAEVADSVEQAIASFGGTVRRVTADPLSTGRDGLRALLA
                PVVAGEEPLTGAVSLLGLCTDGHPEHPGVPAGVTGTLALVQALADLGVAVPLWTVTRG
                TVLVVGDSTTVPARLLRSLLDDGAERVVLAGADAGTATDAGTAELAPDAAVVRIPCDV

  Query Match           53.5%;  Score 285.8;  DB 1;  Length 41097;
  Best Local Similarity  73.4%;  Pred. No. 2.8e-27;
  Matches  394;  Conservative    0;  Mismatches  137;  Indels    6;  Gaps    2;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || |||| |||||||||||||| |||||||||||| |||||| || ||  |   |||
Db      33786 GGCACCATCCTCATCACCGGCGGCACCGGCGCCCTCGGCACCCACATCGCCACCTGGCTC 33845

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
              |||||  ||   | || |||||||| |||||||||||||    || ||| || |||||
Db      33846 G---CCCACAAAGGCGCCAAACACCTCATCCTCACCAGCCGCCAAGGACCCAACGCCCCC 33902

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              |||    || |||     ||| |||   |||   |||| ||||  |||||||||| ||
Db      33903 GGCGCCACCCAACTCCAACACCACCTCACCCAACTCGGCACCCACACCACCATCACCCCC 33962

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||  || | ||||  | ||| || || | ||| |||||  ||||  | | || |||
Db      33963 TGCGACCTCACCGACCCCCAACAAATCGAAAACCTCATCAACCACATCAACACCCACCAC 34022

Qy        241 ---CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAAC 297
                 ||||||    || || ||||||||||| || | || |  |  | || | ||||||  ||
Db      34023 CCACCCCTCCAAGCCATCTTCCACACCGCCGGAACCCCCCAGACCACCCCACTCACCCAC 34082

Qy        298 CTCACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTC 357
              ||||||||||||||  || | || ||||  ||||||| ||  | ||||| | ||||||
Db      34083 CTCACCCCCACCCAAGCCACCGACATCCTCAACGCCAAAACCACCGGCGCCGAACTCCTC 34142

Qy        358 CACCAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCC 417
              ||||| | ||||  ||||||||||| | || |||||| || ||||||||||||||||
Db      34143 CACCACCACACCACCCACAACCCCCCTCCACACCGTCGTCCACTTCTCCTCCGCCGCCGCC 34202

Qy        418 ACCTTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTC 477
              |||| |||  ||| |||| |||||||||| |||||||||| ||||||||||||||||||
Db      34203 ACCTGGGGCAGCAGCGGACAATCCCTCTACGCCGCCGCCAACGCCCACCTCGACGCCCTC 34262

Qy        478 GCCCACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              ||||| |||  |||||||||   ||  ||| |||||  ||||||||||||     |||
Db      34263 GCCCAACACACCCACACCCAAGGCCAACCCACCACCTCCATCGCCTGGGGACTGTGG 34319


RESULT 7
SC2C4/c
LOCUS       SC2C4                  24225 bp    DNA     linear   BCT 12-MAY-2002
DEFINITION  Streptomyces coelicolor cosmid 2C4.
ACCESSION   AL512902 AL645882
VERSION     AL512902.2  GI:20520683
KEYWORDS    accA1; secreted FAD-binding protein; type I polyketide synthase.
SOURCE      Streptomyces coelicolor A3(2).
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 24225)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
   PUBMED   8843436
```

```
Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              |||     | |||    || |||||    |   |||  |||| |||||  ||||| ||
Db      96548 GGCGCCACCGAACTCGTAGCCGAACTCACCGCAGCCGGCGCCCACGTCACCGTCACCGCC 96489

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              |||||||||| | |||| || ||||||       |||||| ||||||||||  |  |||||
Db      96488 TGCGACACCACCGACCCCGAACAACTCGCCACCCTCCTCACCACCATCCCCACCGAACAC 96429

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              |||||||||||||||||||||||||||||| ||||||||||||||   ||| || ||
Db      96428 CCCCTCACCACCGTCATCCACACCGCAGGCACCCTCGACGACGCCACCACCACGAAGCTG 96369

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              |||  |||||||||| ||| ||||||||   ||||   | ||   |||  || |||||||
Db      96368 ACCGACACCCAACTCCACACCGTCCTCCAACCCAAGATCGACGCAGCCACCCACCTCCAC 96309

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              || |||||||  |||  |||| ||||| ||||||||||||||||||||||||||||
Db      96308 CACCTCACCCTCCACCACCCCGTCACCACCTTCGTCCTCTACTCCTCCGCCGCCGGACAA 96249

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              |||||  |  | || |||||||||||||||| |||||| ||||||||||||||||||||
Db      96248 CTCGGCACCGCAGGACAAGCCAACTACGCCGCCGCCAACACCTACCTCGACGCCCTCGCC 96189

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              |||||||||| |||||||   | ||||||||||  | ||||||||||  ||||
Db      96188 CACCACCGCCGCACCCACGGACACCCCGCCACCTCCCTCGCCTGGGGCTTCTGG 96135
```

RESULT 6
AF016585
LOCUS AF016585 41097 bp DNA linear BCT 07-DEC-1997
DEFINITION Streptomyces caelestis cytochrome P-450 hydroxylase homolog (nidi) gene, partial cds; polyketide synthase modules 1 through 7 (nidA) genes, complete cds; and N-methyltransferase homolog gene, partial cds.
ACCESSION AF016585
VERSION AF016585.1 GI:2558836
KEYWORDS .
SOURCE Streptomyces caelestis.
ORGANISM Streptomyces caelestis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 41097)
AUTHORS Kakavas,S.J., Katz,L. and Stassi,D.
TITLE Identification and characterization of the niddamycin polyketide synthase genes from Streptomyces caelestis
JOURNAL J. Bacteriol. 179 (23), 7515-7522 (1997)
MEDLINE 98053867
PUBMED 9393718
REFERENCE 2 (bases 1 to 41097)
AUTHORS Kakavas,S. and Stassi,D.
TITLE Direct Submission
JOURNAL Submitted (30-JUL-1997) 47p, Abbott Laboratories, 100 Abbott Park Road, Abbott Park, IL 60064, USA
FEATURES Location/Qualifiers
source 1. .41097
/organism="Streptomyces caelestis"
/strain="NRRL-2821"
/db_xref="taxon:36816"
gene complement(<1. .396)
/gene="nidi"
CDS complement(<1. .396)
/gene="nidi"
/note="ORF6; similar to tyli"
/codon_start=1
/transl_table=11
/product="cytochrome P-450 hydroxylase homolog"
/protein_id="AAC46023.1"
/db_xref="GI:2558837"
/translation="MVDSVTGPMELSKDANAKELLDWFSHNRTHHPVFWDEGRQAWQV
FRYDDYLTVSNHPEFFSSDFTEVAPTPPELEMILGPGTIGALDPPAHGPMRKLVSQAF
TPRRMAGQEQRIRVIAEELLDRVRGQKTIA"

gene 627. .13649
/gene="nidA1"
CDS 627. .13649
/gene="nidA1"
/note="orf1"
/codon_start=1
/transl_table=11
/product="polyketide synthase modules 1 and 2"
/protein_id="AAC46024.1"
/db_xref="GI:2558838"
/translation="MAGHGDATAQKAQDAEKSEDGSDAIAVIGMSCRFPGAPGTAEFW
QLLSSGADAVVTAADGRRRGTIDAPADFDAAFFGMSPREAAATDPQQRLVLELGWEAL
EDAGIVPESLRGEAASVFVGAMNDDYATLLHRAGAPTDTYTATGLQHSMIANRLSYFL
GLRGPSLVVDTGQSSSLVAVALAVESLRGGTSGIALAGGVNLVLAEEGSAAMERVGAL
SPDGRCHTFDARANGYVRGEGGAIVVLKPLADALADGDRVYCVVRGVATGNDGGGPGL
TVPDRAGQEAVLRAACDQAGVRPADVRFVELHGTGTPAGDPVEAEALGAVYGTGRPAN
EPLLVGSVKTNIGHLEGAAGIAGFVKAALCLHERALPASLNFETPNPAIPLERLRLKV
QTAHAALQPGTGGGPLLAGVSAFGMGGTNCHVVLEETPGGRQPAETGQADACLFSASP
MLLLSARSEQALRAQAARLREHLEDSGADPLDIAYSLATTRTRFEHRAAVPCGDPDRL
SSALAALAAGQTPRGVRIGSTDADGRLALLFTGQGAQHPGMGQELYTTDPHFAAALDE
VCEELQRCGTQNLREVMFTPDQPDLLDRTEYTQPALFALQTALYRTLTARGTQAHLVL
GHSVGEITAAHIAGVLDLPDAARLITARAHVMGQLPHGGAMLSVQAAEHDLDQLAHTH
GVEIAAVNGPTHCVLSGPRTALEETAQHLREQNVRHTWLKVSHAFHSALMDPMLGAFR
DTLNTLNYQPPTIPLISNLTGQIADPNHLCTPDYWIDHARHTVRFADAVQTAHHQGTT
TYLEIGPHPTLTTLLHHTLDNPTTIPTLHRERPEPETLTQAIAAVGVRTDGIDWAVLC
GASRPRRVELPTYAFQRRTHWAPGLTPNHAPADRPAAEPQRAMAVGPVSREALVRLVG
ETTASVLGLDGPDEVALDRPFTSQGLDSMTAVELAGLLGTAAGVALDPTLVYELPTPR
AVADHLAKTLLGESAADADQEVNGRTGEAEAKAGDPIAVIGIGCRFPGGVATPDDLWE
LVASGTDAISTFPTDRGWDLDGLYDPDPSTPGKSYVRHGGFLHDAAQFDAEFFGISPR
EATAMDPQQRLLLETSWEALERAGVVPESLRGGRTGVFVGTTAPEYGPRLHEGTDGYE
GFLLTGTTASVASGRIAYALGTRGPALTVDTACSSSLVALHLAVQSLRRGECDLALAG
GTTVMSGPGMFVEFSRQRGLAPDGRCKAFSADADGTAWAEGVGMLLVERLSDAERLGH
RVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALSDAGLSADDIDAVEAHGTGTAL
GDPIEAGALLATYGHPKRQTPVWLGSLKSNIGHTQAAAGIAGIIKMVQALRHDTLPRT
LHADHPSSKVDWDAGPLQLLTDARPWPADPDRPRRAGISAFGVSGTNAHVVLEEPPAP
EEAPARTQAPPPVIAWPLSAHTPTALRAQAARLDSWLHHTQADPQDIAHALATTRTHF
KHRAVITGRTRTELHTKLHTLDAIQGTAHPHPRLTLLFTGQGAQHRGMGQELYATDPH
FAAALDEVCEELQRCGTQNLREVMFTPDQPDLLDRTEYTQPALFALQTALYRTLTARG
TQAHLVLGHSVGEITAAHIAGVLDLPDAARLITARAHLMGQLPHGGAMLSVQAAEHDL
DQLAHTHGVEIAAVNGPTHCVLSGPRTALEETAQQLHQQGIRHTWLKVSHAFHSALMD
PMLGAFRDTLNTLNYQPPTIPLISNLTGQIADPNHLCTPDYWIDHARHTVRFADAVQT
AHDQRTTTYLEIGAHPTLTTLLHHTLDNPTTIPTLHREHPEPETLTTALATLHTTGHT
TTLHTTSPQTHHLDLPTYPFQRDRYWMEPVRVAQVSGQPGADRLRYRVVWEAAPEDHT
VPSAERWLLLGGSDPADAGILVGVEEQLVGHGAAVQRLDIPAPAGRRLLPSCSELPVR
GPSHTGEFCGSRLRSPRVAEAVALLQALGDAGADAPLWIATRGAVAAQAGEAPSVGGA
QLWGLGQVAGLELADRWGGLVDLPADPAPAALRGLARVLVSNAPDNQVAIRASGVFVR
RVVPAPGRPVRNDWAPSGTVLITGGTGALGSQVARRLALAGAPHLLLAGRRGNSGAAE
LVDELTALGAEVTVAACDAADRDALVSLLATIPEHRPLTAVLHAAGVLDDGVLDGLTP
ERIDAVLRAKATAARHLDELTADLDLDAFVLFSSIVGVWGNGGQATYAAANAALDALA
HGRRARGQRATSIAWGPWAGSGMAAGDGAKSLARDGIDVLEPEEALDLLEQAVRSDET
TLLVADVDWETFLSSSASRRFQLPLFGRIPAARAVAVSGADGSPDDVPVWLSPQLSAA
ERQLRLLDLVRTEAASVLRHGSATAVDAETAFRAGGFDSLTLVELRNRLASATGLRLS
NTLLFDHPTPAAVAEYLGEQLLGTEPSAAASHGDAGARTVVASDEPVAVVGMACRYPG
GVATPDDLWELVASGTDAISTFPTDRGWDLDGLYDPDPSTPGKSYVRHGGFLHDAAQF
DAEFFGISPREATAMDPQQRLLLETSWEALEHAGIAPHTLRTTRTGVFTGASQQEYGT
QSREAADKYGGHLLTGTLASVMSGRVAYTLGLQGPALTVDTACSSSLVALHLAVQSLR
RGECDLALAGGSTVMSTPTVFVEFSRQRGLAPDGRCKPFADTADGTAWAEGVGVLLVE
RLSDAERLGHRVLAVVRGTAVNQDGASNGLTAPSGPAQQQVIRDALSDAGLSADDIDA
VEAHGTGTPLGDPIEAEALLATYGHPERQTPVWLGSLKSNIGHSQAAAGIAGIIKMIQ
AMHHGTLPRTLHVDAPTSQVDWEAGSLQLLAEARPWPADPDRPRRAGISAFGVSGTNA
HVVLEEPPAPEEAPARTQAPPPVIAWPLSAHTPTALRAQAARLDSWLHHTQADPQDIA
HALATTCTHFKHRAVITGRTRTELHTKLHTLDAIQGTAHPHPRLTLLFTGQGAQHPGM
GQELYTTDPHFAAALDEICEELQRCGTQNLREVMFTPDQPDLLDRTEYTQPALFALQT
ALYRTLTAHGTQAHLVLGHSVGEITAAHIAGVLDLPDAARLITARAHLMGQLPHDGAM
LSVQAAEHDLDQLAHTHGVEIAAVNGPTHCVLSGPRTALEETAQHLREQNVRHTWLKV
SHAFHSALMDPMLGAFRDTLNTLNYQPPTIPLISNLTGQIADPNHLCTPDYWIDHARH
TVRFADAVQTAHDQRTTTYLEIGPHPTLTTLLHHTLDNPTTIPTLHREHPEPETLTTA
LATLHTTGHTTTPHPSHIPAQRVSLPAYPFQRRAYWMPNSAAHIGRSDAEAATRLGLA
RQDHPFLTGATSIVGSGATLLTGRVSLAAHPWLADHTVAGAVLFPGTAFADLLLHVAE
ETGSGTVEELTLHVPLVLPAQGAVQLQVLVEAPDERGRRAATVAARPEEETGAEWTTH
AAGVLAGAEPAAPAVDWADWAAGAWPPPGAEAVDVDELYAEFAAAGYDYGPAFTGLTG
AWRRGEEVFADVRLPEAVAGSGEEERRFGVHPALFDAALHPWRVGGPWQESGPDADGG
QSTLLPFSWQGLTLYGTGAEQVRVRLAPSDGGAFSVRAADPAGHPVLALDALVLRRVP
```
```

```
                /protein_id="BAB69187.1"
                /db_xref="GI:15823970"
                /translation="MSQDRHDGSTAALGGPSHGGRLLAVSDIHVRHRQNRDIVQDLHP
                GCAEDWLIVAGDVAESVADITWALSLLRDRFAKVIWAPGNHELWTAPGDPDELRGERK
                YRHLVDLCRALGVVTPEDPYPVWHGAGGPVAVAPLFLLYDYSFLPPGAWSKSHALRLA
                DDAGVICTDEFLLHPDPYPSREAWCRARVRATRARLDALPDDLPTVLVNHFPLVREPT
                RVLRHPEFALWCGTTATADWHLRYRAAAVVHGHLHIPRTIRCDDVPHHEVSLGYPREW
                QRYGRRPDWPVQILTPAPAPAPAPVPVPVPAPPGHRTASPADRRP"
gene            3213. .3971
                /gene="olmC"
CDS             3213. .3971
                /gene="olmC"
                /codon_start=1
                /transl_table=11
                /product="thioesterase"
                /protein_id="BAB69188.1"
                /db_xref="GI:15823971"
                /translation="MSANRVDGNLWVRQFHPAPTAATKLVCLPHAGGSASYFLPVSKA
                LAPGVDVLAIQYPGRQDRRHEKCIDNIAELADAVTAVLEPLADRPLVLFGHSMGATLA
                FEVALRLERKGIVPLALFASGRRAPSRHRDESVHLRGDDEIIAELKTLSGTRSEVLGD
                EEILRMVLPAIRSDYKAAETYRYAGGARLTAPIHAHFGVEDPRVSLDEAQAWADHTTG
                HFELRSHPGGHFYLNDQAPRVIEEIARVSSSVAS"
gene            4643. .7438
                /gene="olmRI"
CDS             4643. .7438
                /gene="olmRI"
                /codon_start=1
                /transl_table=11
                /product="luxR-family transcriptional regulatory protein"
                /protein_id="BAB69189.1"
                /db_xref="GI:15823972"
                /translation="MLIGRGQEIDRLRAAYSSVRSTGRSRAVLVESAAGCGKSEILED
                LVRHVESLGLIALRATAIAEERNTPLSVLRSLVNGAPFPDDLVRRFHTMSDAAARMIS
                TEGRDKARRDAAWYELMRTFCTEVRALAQRAPLVVAVDDIHHGDADSLRQLLHLARHC
                RNVPVLLTFTQSLAGGPADRTFGAELLRQPNLEHIRLPRFCRQDIADMLAEQGRADAG
                QGTVERYFAVSGGNPLLLRALLEDARLEPDDTVDAQDVGPRAGQAYGRAVLACLDRGG
                EKAHEVAAGLAVLGGAFTHELLSRLLDISAAALTENLRALRAAGIVVGTSFQHPAAKA
                AVLDSLEPSHRLRLHQRSAALLHASGAPAALAAEHLHEAGQADGAWAVGVLQQAAEQA
                LLADEARRAVSYLELAHAASADPRGRTEIKIRLGSVTRRVSAADAERHVDDALRGFGA
                GLLTPSDTGGLAALLLAHGRLEQARSVLEQGRRPGDGDADRTAARRQLAEIRDSVLNR
                GAWCAGTRAAWHENPAGRPSTGRAGTGLRAWLSGSHAALLAQRIAGTEGLLEITHLTD
                ATFDLVLNAVWTLLALGAVERALHWCDTFLREAKRRDAPGWEAVFAAVRAEGALRGGN
                LAEAERDARRCFAVLPERHRSALEGSAVSYLVTVQIARGKYEDAARQLSRPMADSLPD
                TIHWLDYLRARGLYYMATRQYHSALRDFHEVGRLAERWGTDWPALLPWRTDTAEALLR
                LGEREQGLRLATEQLAMVQDGAPRIKGISLRLQGLASEPAARLGLLNRAVVELSRADD
                RLELSRTLFDLADAHRETRQTRQAAMVLRRARQLAKSCGAQPPRDRIGSPDAAGASGE
                LFGEGPERNADWGHLSESERRVAGLAASGYTNREISVRLYITMSTVEQHLTRVYRKLK
                INGREELLAEARAGVQV"
gene            7658. .10486
                /gene="olmRII"
CDS             7658. .10486
                /gene="olmRII"
                /codon_start=1
                /transl_table=11
                /product="luxR-family transcriptional regulatory protein"
                /protein_id="BAB69190.1"
                /db_xref="GI:15823973"
                /translation="MLASAGQGEGGVVLVNAGPGVGRTALLDAFLRKATESGARVCAA
                AGSSAETEAEFGVVGQLFPPTGPLGAGFRLARAIGSRDAGPGLDADRLFDALCQELRR
                RPTVIAVDDAQLADAASLRFLRYLVRRLRTAPVMMVLTEPTGPDALPPSFQAELLRHP
                QCRRLGLLPLSRDGVTQVVSAQVGSDRVPRLAADFHAISGGNPLLVHALLADHRAAHR
                LSPVPSVPTVPSVSSASSAASASSAASAASAVSAAPAASAAFAKAALAWAYRDEPVLF
                DVACGFAVLGESATPARVACLVDRGADAVDQVMTALDTAGLLAGGTFRSALVGKALLE
                HVDVETRAELRRRAARLLHGDGAPATVVARHLLAAPVAEPWAGRTLLAAAEKAMQEGR
                EELSLDCLRLAGQSATTEQERAAVVMARVKIGWEIDPRLVSPWLDELVAALRRGDLRS
                GDAAWIVKHLAWQDRFPEAVESLSTLAAAGDAGEPGVEAELHFVRHWLRHSCPPLEAS
                VPGPSNPTRPRTPRLRRVSPASYAGEVLGKVLTEGPDDHTVAMAEEILRGCRYSETVV
                EALEAALLALVYAGRPDRAAFWCDGLLQQAGTRSTGTSVAILSGIRAEIALRQGALPA
                VEQYARSALSAISLPGWGVAVGSPLAALVHSTTASGMSGVAGAWLNQAVPDGMFRTRH
                GLLYTHARGHYHLAVDRPAAALDDFLTCGGLAKEWGMDVPTFLPWRTSAAQAQLALGN
                PGQARALVREQLGRPGGSTARARGVALRLLAATGEPGRRPALLRESVNLLESCGDQVE
                LLRALADRCQALDHVGAPAKARMTARHARNVADACGGDALFRKLFRDEVPDREGDPDG
                LTGEEQGSASLTAAERRVTALAALGHSNREIGRKLFITKSTVEQHLTRVYRKLGVRNR
                ADLGDLLAGTNLGAGTPAMAMASSAVAG"
CDS             10933. .11271
                /codon_start=1
                /transl_table=11
                /product="P450-like protein"
                /protein_id="BAB69191.1"
                /db_xref="GI:15823974"
                /translation="MEVGGQTIRAGEGVLCTLSTANRDEGAFTDPDELDLRRDARSHL
                SFGFGIHQCIGQSLARAEPQIALETLLRRLPGLRLAASFEELRFRDGVGFYGVKELPD
                PCPALPPDRP"
gene            11458. .29898
                /gene="olmA1"
CDS             11458. .29898
                /gene="olmA1"
                /codon_start=1
                /transl_table=11
                /product="modular polyketide synthase"
                /protein_id="BAB69192.1"
                /db_xref="GI:15823975"
                /translation="MACRLPGAAGPAEFLALLRNGEDAVGAAPGDRRYAPRRGGFLDS
                VDEFDAGFFGISPREAAAMDPQQRLVLELSWEALEDAGIPPARLGGSGAGVFVGAIAD
                DYATLSRAAGADAATPQTTTGLNRGIIANRVSYTLGLRGPSFTVDSGQSSSLVAVHLA
                AESLRRGECSVALAGGVNLILAEDSTVAVERFGALSPDGYCYTFDARANGYVRGEGGG
                VVVLKPLADAIADGDDIHCILAGSAVNNDGGGEGLTVPDRAGQEAVLTAAYEQAGISP
                AAVGYVELHGTGTPTGDPVEAAAVGAVLGAGRPAGQPLLVGSVKTNIGHLEGAAGIAG
                LLKAVLAVRHREIPATLHHASPNPAIPMAELGIGVNTALTPWPEQAGPLVVGVSSFGM
                GGTNCHVVLTEPAGYDAAARPATAGIRPAGTPAPLLVSGRDEQALRAQARKLRDHLDT
                HPGLPLRDVAHTLAAARAPLARRAVLLAADREDVSAGLARLADGTPGADLVRGTVSPG
                ALAFVFTGQGSQRPGMTAELYRTSAEYAAALDEVCASLDPWLRRPLRDVLFAAEGSDE
                AALLDRTEFTQPALFAVEVALFRYAEHCGLTPRFLLGHSVGELAAAHVAGVLSLADAC
                TLVAARGRLMQDRPATGAMVSVQATEDELAPFVSESVTLAALNGPASTVIAGDEDAVL
                AVAAHWAARGRKTKRLRVSHAFHSPHMDGMLTEFRRIAAGLTFGAPRIPIVSNETGTL
                LTEADARSPEYWVRHARAAVRFLDGVRLLEDQGVTTLLELGPDGVLSAMARDCLRDAG
                TVCVPLLRGRDRTEPEALLAALAGLYVRGVPMDWERTATEAGARRVPLPTYAFQRRRH
                WLPEPSASASRPVVNGGGHAGQAGHAGQAGRAGQAGHAGHDGEPAPAVRDRTTPELVR
                DRTTLELVRDATAQVLGHASPDDVATHEAFRELGFSSLMLAELGERLTEATGRRVPTT
                LLFDHPTPDALARRLEADGAERTDAGPVSPPAAAAHDDDPIVIVGMGCRLPGGVRSPE
                DLWRFVAAGGDAISALPVDRGWEFAADFPGEGGFLADVAEFDAGFFGISPREALAMDP
                QQRLLLETSWEALERAGVDALKLRGSRTGVFVGASPSEYGPRLHETSDADGHVLTGTA
                PSVLSGRIAYVLGLEGPALTVDTACSSSLVALHLAAQALRGGECDLALAGGVTVMATS
                GMFAEFARQGGLAPDGRCKAFADGADGTGWGEGVGMLAVQRLSDAVRDGRPVLAVVRG
                SAVNSDGASNGLTAPNGPSQQRVIRQALASAGLSAADVDMVEAHGTGTSLGDPIEAQA
                LLATYGQDRPDDRPLWLGSVKSNIGHTQAAAGVAGVIKSVLALRNGLLPKTLHVDEPT
                PEVDWSAGAVELLTEGREWPETDGPRRAGVSAFGISGTNAHVILEQAPSVEEPVTAPA
                DDGLLVPWVVSARSEEALRAQARRLADHVRESGLRPADVGLSLAVTRAGLEHRAVLVA
                EDEAGFLAHLDALAEGRTPAGVVVEGGTAFLFSGQGSQRAGMGRELYSAHPVFAAAFD
                EVCAAFDGLLPGALREVVFGGAGLDRTEWAQPALFALEVALFELVSSWGVRADVLLGH
                SIGELAAAYVAGVWSLADACRIVAARGRLMQALPEGGAMVAVEAAEDELPDLPGGVSV
                AAVNGPRSLVLSGDEEPVTALAQAFAEQGRRTKQLTVSHAFHSARMEPMLAEFAETLA
                AVEFRPPRIPVVSNVTGEMAGDEFTTPAYWVRHVREAVRFADGVGAAHAAGVRRYLEL
                GPDGTLTTLAHDTAHDAGRDVAAVPALRGDQPETRTVVTAVGRLYATGLTVDWPAFFA
                PYGARPTELPTYAFQHRRYWLEPVAPSSAPPANPLRYRTAWASVPDGPRPSLSGTWAV
                VVPASGSYEDIVAGVTAALRDHGARPVVVEVDPAGPERADVADTLRAALGEDAGGVVS
                LLGTAGSLSSGVAATAVLVQAWTDLGGGARLWCLTRGAVSVSPSDPLAAPAQAQLWGM
                GRVAALEHPERWGGLVDLPAVPDDRAWTRLCAVLAGGSGEDQVAVRAAGLFARRLGRA
                GTRADAAPAGPSPWRTDGTVLVTGGTGALGAHLAQWLAAAGAAHVLLTSRRGPDADGA
                AELTARLRETGTEVTVAACDVTDRDALAELLSGLPADRPLTGVVHAAGVLDDGVLDSL
                TPDRFAAVAGPKVTGAWHLHELTRDLDLSAFVLFSSFAGTVGLAGQANYAAANAYLDA
                LAVHRAQLGLPATAVAWGSWSGAGMAADTEAARRQLARTGLVPLAPDAALAALGQVMD
                DGETAVTVADVDWDRFAAGFGPGRPSPLLTGVPEVRRIEERETGRAGAPAETDRLGGL
```

Query Match 62.8%; Score 335.6; DB 1; Length 104326;
Best Local Similarity 76.8%; Pred. No. 1.5e-33;
Matches 410; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

```
Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              || ||| ||||||||||||||||| |||||| |||||| | ||| |||| ||| |||||||
Db      96668 GGCACCGCCCTCATCACCGGCGGCACCGGCACCCTCGGCGCCCTCCTCGCCCGCCACCTC 96609

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
                |||||  ||              ||| || |||||||||||| ||| ||||| |||||
Db      96608 GTCACCCGGCACGGGGTGACCAGCCTGCTGCTCACCAGCCGACGCGGACCCCACGCCCCC 96549
```

```
                     TQLQQKGIHLTITTCDTSTPRPTHNNSLNTIPPQHPVTTVIHTGGILDDATLTNLTPT
                     QLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAH
                     HRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIADTER
                     PYVVIADIDWSKIEHTSQTSDLVSAARAEREPAVQRPTPPAELHKTLAHQTSADQRAAL
                     LELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFD
                     HPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDF
                     WDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGIS
                     PREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQS
                     TDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMAL
                     AGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRN
                     GHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGT
                     TLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLL
                     PRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEA
                     PADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPG
                     LDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGT
                     GEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQN
                     DNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITGAHLAG
                     ILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSL
                     VISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHT
                     PLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHH
                     NLPNTPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHAQPTPHPHPLDLPTYP
                     FQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLA
                     DHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAAD
                     GKGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAV
                     AVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPAL
                     LNGVVQPLSVLLPGGTGFGKGAGFGKGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGG
                     GREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVP
                     VCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQG
                     WLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDT
                     DTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLTRLESSGDVPAQRS
                     GDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLA
                     GVCGVRDLLLVSRRGPDAPGAEGLRRSWPRGAEVRIVACDVGERREVVRLAGGCSCRV
                     SVFVDFSAVASVTVRLPVASDVRKEAAMAYATVEEFTDYLDPDP"
BASE COUNT     1756 a   4059 c   3528 g   1753 t
ORIGIN

  Query Match           93.2%;  Score 497.8;  DB 1;  Length 11096;
  Best Local Similarity   99.1%;  Pred. No. 1.6e-53;
  Matches  532;  Conservative    0;  Mismatches    2;  Indels    3;  Gaps    3;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5260 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 5319

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5320 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 5379

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5380 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 5439

Qy        181 TGCGACACCAGCA-ACCCAGACCAACTC-CAACAACTC-CTCAACACCATCCCCCCACAA 237
              ||||||||||||| ||||||||||||| ||||||||| |||||||||||||||||||||
Db       5440 TGCGACACCAGCACACCCAGACCAACTCACAACAACTCACTCAACACCATCCCCCCACAA 5499

Qy        238 CACCCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAAC 297
              |||||| |||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       5500 CACCCCGTCACCACCGTCATCCACACCGGAGGCATCCTCGACGACGCCACCCTCACCAAC 5559

Qy        298 CTCACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTC 357
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5560 CTCACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTC 5619

Qy        358 CACCAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCC 417
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5620 CACCAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCC 5679

Qy        418 ACCTTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTC 477
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5680 ACCTTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTC 5739

Qy        478 GCCCACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5740 GCCCACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 5796


RESULT 5
AB070940/c
LOCUS       AB070940              104326 bp    DNA     linear   BCT 22-JAN-2002
DEFINITION  Streptomyces avermitilis oligomycin biosynthetic gene cluster.
ACCESSION   AB070940
VERSION     AB070940.1  GI:15823967
KEYWORDS    .
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
  TITLE     Genome sequence of an industrial microorganism Streptomyces
            avermitilis: deducing the ability of producing secondary
            metabolites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
  MEDLINE   21477403
REFERENCE   2  (bases 1 to 104326)
  AUTHORS   Ikeda,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .104326
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            complement(1. .1134)
                     /gene="mcmA"
     CDS             complement(1. .1134)
                     /gene="mcmA"
                     /note="methylmalonyl-CoA mutase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB69185.1"
                     /db_xref="GI:15823968"
                     /translation="MTGYSTAADANARFRQLVGSGTRAVPVVFDLPTRAGLDSDAPAA
                     RGLVGRRGVAVDSIDDMRVLFGGLRLDRVVPSLAVDAPAAPLLVLYQLVAEEHGAPAQ
                     RLAGAVRGDVFKEFLARGPGVFPLRPSLRLAMNVFAYCQAELPRWQCLSVCGHHLAGT
                     GADPALEVAFTVAGGIEYLRAAVTAGLDVRTVAPRLTLCLVGGTTGRENRAKLRATRR
                     VWTRTANERFGSGRTAPRLYVRTPRDGLATLACTEDIEAGATALVDRIERLGGTLTVL
                     EHGLAGQLLGRPHGDRPATPRPHPPHQDSLDLQARQTERLAKLRAWRVQPAVDEALLR
                     LRKAAQGDDNVLYPMREALAARATVGEVCDALRSTWGTGPSAT"
     gene            complement(1131. .1886)
                     /gene="pptA"
     CDS             complement(1131. .1886)
                     /gene="pptA"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphopantheteinyl transferase"
                     /protein_id="BAB69186.1"
                     /db_xref="GI:15823969"
                     /translation="MRAHEVFADVPVSSLHPREAALVQDAVESRRREFATGRHCARRA
                     LASLGVPPGPLLATPRGAPRWPAGVCGAITHCAGYRAAAVARAPGIRAVGIDAEPDRP
                     LPEGVLEAVALPAEQFQVKEHTARHPGVVWGKLLFSAKESVYKAWFPATGRPLGFEEA
                     LIDLDPVSSTFRVRVLPTALRAAEPELRDLHGRWLVRHGIIVTAVVVDGPGRDTPHAP
                     SPYDRTKHSAPTGRSADRPGGRPTDRATDRPTA"
     gene            complement(1910. .2875)
                     /gene="simX4 homolog"
     CDS             complement(1910. .2875)
                     /gene="simX4 homolog"
                     /codon_start=1
                     /transl_table=11
                     /product="SimX4 homolog"
```

```
Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5603 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 5662

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5663 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 5722

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5723 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 5776


RESULT 3
AX006889
LOCUS       AX006889               12381 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 1 from Patent WO0001827.
ACCESSION   AX006889
VERSION     AX006889.1  GI:9994904
KEYWORDS    .
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 12381)
  AUTHORS   Kellenberger,J.L., Leadlay,P.F., Staunton,J., Mcarthur,H.A. and
            Stutzman-Engwall,K.J.
  TITLE     Polyketides, their preparation, and materials for use therein
  JOURNAL   Patent: WO 0001827-A 1 13-JAN-2000;
            KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
            STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PFIZER (US); MCARTHUR
            HAMISH ALASTAIR IRVIN (US); STUTZMAN ENGWALL KIM JONELLE (US)
FEATURES             Location/Qualifiers
     source          1. .12381
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     1884 a   4561 c   4005 g   1931 t
ORIGIN

  Query Match             99.7%;  Score 532.4;  DB 6;  Length 12381;
  Best Local Similarity   99.8%;  Pred. No. 7.8e-58;
  Matches  533;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5960 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 6019

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6020 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 6079

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6080 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 6139

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6140 TGCGACACCAGCAACCCAGACCAACTCCAACATCTCCTCAACACCATCCCCCCACAACAC 6199

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6200 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 6259

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6260 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 6319

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6320 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 6379

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6380 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 6439

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6440 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 6493


RESULT 4
AF275943
LOCUS       AF275943               11096 bp    DNA     linear   BCT 02-SEP-2000
DEFINITION  Streptomyces avermitilis avermectin polyketide synthase gene,
            partial cds.
ACCESSION   AF275943
VERSION     AF275943.1  GI:9964075
KEYWORDS    .
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 11096)
  AUTHORS   Hong,Y.-S. and Lee,J.J.
  TITLE     Targeted Gene Disruption of the avermectin O-methyltransferase gene
            and polyketide synthase gene from Streptomyces avermitilis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11096)
  AUTHORS   Hong,Y.-S. and Lee,J.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2000) Anticancer Agent Research Laboratory, Korea
            Research Institute of Bioscience and Biotechnology, P.O. Box 116,
            YuSong-Gu, Taejon 305-600, South Korea
FEATURES             Location/Qualifiers
     source          1. .11096
                     /organism="Streptomyces avermitilis"
                     /strain="ATCC31271"
                     /db_xref="ATCC:31271"
                     /db_xref="taxon:33903"
     CDS             220. .>11096
                     /note="Avr1"
                     /codon_start=1
                     /transl_table=11
                     /product="avermectin polyketide synthase"
                     /protein_id="AAG09812.1"
                     /db_xref="GI:9964076"
                     /translation="MADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAE
                     FAPYVDWSVEQVLRDSPDAPGLDRVDVVQRTLFAVMISLAALWRSQGVEPCAVLGHSL
                     GEIAAAHVSGGLSLADAARVGDAWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDN
                     PARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEER
                     VVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGP
                     KTFVEMSPHPVLTMGLQELAADLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGH
                     GETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHL
                     GFDSVMGVELRNRLSKATGLRLPVTLILDHTTPAAVAARLRTAALGHLDEDTAPVPDS
                     PSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQR
                     HAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERA
                     GIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRSAYS
                     FGLEGPAITVDTGCSASLVTLHLACQSLRSGECTFALAGGVSDVHPGMFIEFSRQCGL
                     SVDGRCKAVSAAADGTGWGEGIGLLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQAL
                     ANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMA
                     AAGVGGVIKMVMALREGVLPRTLHVDEPSPQGLAAGAVRLLTEAVPWPGDAAGRLRRA
                     GVSSFGIGGTNAHVILEEAPAAGGCVAGGRVLEGAPGLAISVAESVAAPVAVSAPVAE
                     SVPVPVPVPVPVPVSARSEAGLRAQAEALRQYEAVQPDVSLADVGAGLACRQAVLEHH
                     VVILAACTSSSRSAARTTARSSSTARPQARPAPTSARRIGALAAGSGSAALTTGHAPG
                     GDRGGVVFVFPWQGGQWAGMGVRLLCLLRVFARRMQACEEALAPWVDWSVVDILRRDA
                     GDAVWEQADVVQPVLFSVMVSLAALWRSYGIEPNEVLGHSKDEIAAAHIYGALSLKDA
                     AKTVALPPQEVEQLIGERGGRLWVAAVNGPRSTAVSGDAEAVVEVLAYCAGTGVRPRI
                     PVDYASHCPHVQPLREELLELLGDISPQPYGVPFFSTVEGTWLDTTTLDAAYWYRNLH
                     QPVRFSHDVQALADQGHRVLLEVSPHPTLVPAIEDTTEDTARRRHCDRQPPPRRERHP
                     LLPQRLRLDPYYRHRQTHHVAPLLDPPRHLPPPLDAPRPAHLPFQHQHYWLESSQPGA
                     GSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVV
                     PALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNIL
                     TNLHHHGITPIPLPLTTHHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPH
                     PHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLAR
                     TTLLEHPTHTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTL
                     TPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLT
```

NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
MTTLHTTPHHITHHLTAHENDLAIAAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
NLAKTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY
VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGG
VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
LRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPE
DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
VSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
SLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLGLS"

gene 12071. .30790
/gene="aveA2"

CDS 12071. .30790
/gene="aveA2"
/note="contains Ave module 3, module 4, module 5 and module 6
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 2"
/protein_id="BAA84475.1"
/db_xref="GI:5902892"
/translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGINP
EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
ACQALRSGECTMALAGGASVMATPFVFTEFSRQRGLAADGRCKAFSAAADGTGWSEGV
GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
NDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALHRL
LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
TQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
NRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLTAVPADEPIAIV
GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
LTYHDYAARFPTAPAGFEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
HLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSAAADGTGWGE
GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAD
LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAADDVPGEAAGDDAGTGGEATGPA
AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH
GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
QGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
PQTHLLTNLAKTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
QHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG
TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
GTWQGNGLATGQVSEHLRRRGMFAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA
AVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ
FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI
GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAAL
LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
TLTHPHHHPQTHLLTNLAKTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
KATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD
IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL

Query Match 100.0%; Score 534; DB 1; Length 64957;
Best Local Similarity 100.0%; Pred. No. 3.6e-58;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5243 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 5302

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5303 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 5362

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5363 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 5422

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5423 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 5482

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5483 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 5542

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5543 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 5602
```

```
            THE KITASATO INSTITUTE
COMMENT     OS   Streptomyces avermitilis
            PN   JP 2000245457-A/1
            PD   12-SEP-2000
            PF   24-FEB-1999 JP 1999046961
            PR
            PI   SATOSHI OMURA,HARUO IKEDA
            PC   C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC   (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(11916)
            FT   CDS             (11971). .(30687).
FEATURES             Location/Qualifiers
     source          1. .30690
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     5355 a  12455 c   8617 g   4263 t
ORIGIN

  Query Match           100.0%;  Score 534;  DB 6;  Length 30690;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-58;
  Matches  534;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5143 GGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTC 5202

Qy         61 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5203 ACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCC 5262

Qy        121 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5263 CACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACC 5322

Qy        181 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5323 TGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACAC 5382

Qy        241 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5383 CCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTC 5442

Qy        301 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5443 ACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCAC 5502

Qy        361 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5503 CAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACC 5562

Qy        421 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5563 TTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCC 5622

Qy        481 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5623 CACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGGGGCACCTGG 5676


RESULT 2
AB032367
LOCUS       AB032367               64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
            Sciences, Kitasato University, Microbial chemistry; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .64957
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            101. .12019
                     /gene="aveA1"
     CDS             101. .12019
                     /gene="aveA1"
                     /note="contains a loading module, Ave module 1 and module
                     2
                     multifunctional polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="type I polyketide synthase AVES 1"
                     /protein_id="BAA84474.1"
                     /db_xref="GI:5902891"
                     /translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
                     ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
                     AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
                     VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGARBAVADLVADLTAAQVRTR
                     MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
                     MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
                     RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
                     DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
                     LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
                     GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
                     MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
                     GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
                     GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
                     RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
                     GDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
                     LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
                     AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAG
                     LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
                     EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
                     VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
                     AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
                     AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
                     ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
                     PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
                     THHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
                     EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
                     WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
                     LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
                     PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
                     LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGAL
                     ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
                     QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
                     TAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
                     DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSA
                     ARELREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
                     FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
                     AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
                     YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
                     GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
                     GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
                     GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 1035.85 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_5143_5676
Perfect score: 534
Sequence: 1 ggaaccaccctcatcaccgg.....gcatcgcctggggcacctgg 534

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		534	100.0	30690	6	E38020	E38020 Avermectin
2		534	100.0	64957	1	AB032367	AB032367 Streptomy
3		532.4	99.7	12381	6	AX006889	AX006889 Sequence
4		497.8	93.2	11096	1	AF275943	AF275943 Streptomy
5	c	335.6	62.8	104326	1	AB070940	AB070940 Streptomy
6		285.8	53.5	41097	1	AF016585	AF016585 Streptomy
7	c	266.8	50.0	24225	1	SC2C4	AL512902 Streptomy
8		266.8	50.0	123580	1	AF263912	AF263912 Streptomy
9		266.8	50.0	125401	6	AX211739	AX211739 Sequence
10		260.4	48.8	65140	6	AX211705	AX211705 Sequence
11		257.4	48.2	19791	1	SPFKBAD	Y10438 Streptomyce
12		252.6	47.3	28732	1	AF082100	AF082100 Streptomy
13		246.8	46.2	32870	1	AF007101	AF007101 Streptomy
14		244.4	45.8	113193	1	AF357202	AF357202 Streptomy
15		244	45.7	77534	1	AF235504	AF235504 Streptomy
16	c	242.4	45.4	77534	1	AF235504	AF235504 Streptomy
17		236.4	44.3	30000	6	AX250262	AX250262 Sequence
18	c	234.8	44.0	84985	1	SNA278573	AJ278573 Streptomy
19		226.8	42.5	20394	1	SNA132222	AJ132222 Streptomy
20		226.8	42.5	20394	6	AX067996	AX067996 Sequence
21		225.2	42.2	27541	6	AX211706	AX211706 Sequence
22		225.2	42.2	104326	1	AB070940	AB070940 Streptomy
23		223.6	41.9	37948	1	AF079138	AF079138 Streptomy
24		217.4	40.7	31422	6	E38021	E38021 Avermectin
25	c	217.4	40.7	64957	1	AB032367	AB032367 Streptomy
26		214	40.1	3573	1	AB016763	AB016763 Streptomy
27	c	212.4	39.8	26195	1	SC1G7	AL591083 Streptomy
28	c	210.8	39.5	30000	6	AX250263	AX250263 Sequence
29		210.8	39.5	43280	1	SFU78289	U78289 Streptomyce
30		207.6	38.9	49377	6	I88042	I88042 Sequence 1
31		207.6	38.9	67523	1	SCU24241	U24241 Sorangium c
32		201.2	37.7	47981	1	AF263245	AF263245 Micromono
33		201.2	37.7	47981	6	AX112026	AX112026 Sequence
34		200.8	37.6	107379	1	SHGCPIR	X86780 S.hygroscop
35	c	199.6	37.4	39314	1	SGR300302	AJ300302 Streptomy
36		196.4	36.8	28958	6	AR044578	AR044578 Sequence
37		196.4	36.8	28958	6	I47768	I47768 Sequence 6
38		196.4	36.8	28958	6	I50958	I50958 Sequence 6
39		196.4	36.8	28958	6	I63356	I63356 Sequence 6
40		196.4	36.8	28958	6	I70387	I70387 Sequence 6
41		196.4	36.8	28958	6	I85639	I85639 Sequence 6
42		196.4	36.8	28958	6	I88045	I88045 Sequence 4
43		196.4	36.8	28958	6	I90322	I90322 Sequence 6
44		196.4	36.8	41782	1	AY118081	AY118081 Streptomy
45	c	195.2	36.6	82746	1	AF453501	AF453501 Actinosyn

ALIGNMENTS

RESULT 1
E38020
LOCUS E38020 30690 bp . DNA linear PAT 31-JAN-2002
DEFINITION Avermectin aglycon synthase gene.
ACCESSION E38020
VERSION E38020.1 GI:18626909
KEYWORDS JP 2000245457-A/1.
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 30690)
AUTHORS Omura,S. and Ikeda,H.
TITLE Avermectin aglycon synthase gene
JOURNAL Patent: JP 2000245457-A 1 12-SEP-2000;

THIS PAGE BLANK (USPTO)

```
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 4832  row: O  column: 2
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 918.
FEATURES             Location/Qualifiers
     source          1. .918
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=4832 Col=2 Row=O"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT    16 a     93 c    718 g     60 t     31 others
ORIGIN

  Query Match             8.0%;  Score 73.6;  DB 17;  Length 918;
  Best Local Similarity   45.2%;  Pred. No. 0.00065;
  Matches  287;  Conservative    0;  Mismatches  345;  Indels    3;  Gaps    1;

Qy          2 TCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCT 61
              |||| | | ||   ||   ||| | |  | ||| | | |  ||| | | | |    ||
Db        262 TCTTTGGGNTTGGGGGGGGGGGGGCGGGGNGGGGGGCGGGGGGGGGGGGGCGCGGCGGGG 321

Qy         62 CCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGG 121
                |     || |   |   || ||  |   ||| ||| || ||  | || || | ||| ||
Db        322 GCGGNGGGGGGGGGGGGGGGGGGGGGGCCGCGGGCGGGGGGGGGCGGGGGCGGGGGGGGG 381

Qy        122 ACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCG 181
               || | | || ||       | |  ||| |  |||||  | || | ||| |  ||  |
Db        382 GGCGGGCGGGGGGGGGGGGGGGGGGGGGGGCNGGGGGGGGGGGGGGGGGGGGGGNGGGGG 441

Qy        182 ATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCT 241
              | ||   ||| | |  |    | | |  || |     ||  |    | || |
Db        442 GGGGGGG---GCCCGGGGGGGGGGGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGNG 498

Qy        242 ACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATG 301
              ||     |      | | ||     || |   | | ||| | |   | ||  | |
Db        499 GGGGGGGGGGGGGGGNGGGGGGGGGGGGCCGGCGCGGGGGGGGGCGGGGGGGGGGGGGC 558

Qy        302 TGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGC 361
              | | |||| |  | | | |  ||  |||| |  |   |  |||  | |  |||||| ||
Db        559 GGNGGGGGGGGGGGGGCGGGCGGGGGCGGGGGGGGGGGGNGCGGGGGGGGCCGGGGGC 618

Qy        362 TGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTGG 421
              |  | | |  |||  || | ||   ||      | |  |  | | |||  || || ||
Db        619 GCGGGGGGGGGGCGGGGGGGCGGGGGGGGGGGGGGGGGGGGGCGGCGGCCGGGGGGGCGG 678

Qy        422 AGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCCC 481
              |          || | | || ||| ||||  |   | |||  || ||  |     |    ||
Db        679 GGGGCGGGGGCGGCGGGGGGGGGGGGGGGGGGCGGGGGGGCGGGGGGCGCGGCGGCGGCC 738

Qy        482 GCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCCG 541
              |      |  |     ||||  |  |    ||  | |  || |  || |    | |  |
Db        739 GGGGGGGGGGGGGGGGCGGGGGGGGGGCNGGGGCGGCCGGGGGGGGGGGGGGGGGGGGC 798

Qy        542 GCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTGC 601
```

```
                |    |||| | |||   || ||   |  |  |     |    |    |  |  | |
Db        799 GGCGGGGGGGGGGGGNGGGGGGGGGCGCGGGGGGGGCGGGGGGGGGGGGGGGCGCGGGG 858

Qy        602 AGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGG 636
                | |   ||||| || |   |  || || |  ||||
Db        859 GGGCGGGGCGGGGGGGGGGGGGGGGGCGGGGGGG 893

Search completed: June 18, 2003, 01:00:14
Job time : 1183.72 secs
```

```
  Best Local Similarity   47.9%;  Pred. No. 0.00055;
  Matches  268;  Conservative    0;  Mismatches  289;  Indels    3;  Gaps    2;

Qy         14 CCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCCTCTCCGGTGT 73
              |||    | | || |    | ||   || | | |    |       | | | | || |
Db        374 CGGGGGGGCGGGCGCGGGNCGGNGGCGCGGCGGGGGGGGCGGGTTTGTGCGCCCNGGGGG 433

Qy         74 TCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGACTGGTCTGTGG 133
                |   |  ||  || || | ||| |||||    ||||    |||  |   ||     ||
Db        434 GGGGGGGCGGGGCGCGGGGGGGCGGGGAGGGGGGGGCGGGCGGGGGTGCGAGGGGGTCGG 493

Qy        134 TGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGATGTGGTCCAGC 193
               |||       | |  ||   | |||||  | || | ||| | |||  |  | ||    |
Db        494 GGGAGGAGGGGGGGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGCGGGGGGG 553

Qy        194 CTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTACGGTATCGAAC 253
              | | | | | | |    || |      ||  ||   | || |  |   |||     |
Db        554 CGGGGGGGGGCGGGGCGGGGGGGGGGGGGGGGGGCGGGGGGGGGGGCGCGCGGGGCGGCGG 613

Qy        254 CCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTGTGTGGGGC-- 311
              |  | |     ||   | |  || || |    |||||  ||     | | |||||
Db        614 GGGGGGGCGGGGGGGGGGGGTGCGGGGCCGGGCGGCCGGCCCCGGGGCGGGGGGGGGCGG 673

Qy        312 GCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTGGCCGCTGT 371
              || | |   |  ||   |||    ||  |  |||     | ||| ||| |  |||  |
Db        674 GCGGGGGGGGGCGGGGACGGGTCGGAGGGGGGCGGGCGGGAGCGGGGGGGGGGCGGGGTC 733

Qy        372 GCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTGGAGCAGCTCAT 431
               |||| ||||  || | | | |   | | |  | | | | || || | |  |  |
Db        734 CGGGGGGCGGGCGGGGAGGGGCGGGGCGGCCGGGGCGGGCGGGGGGGGGGCGGGGGCGG 793

Qy        432 TGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCCCGCTCCACCGC 491
              || | |  | ||  |||| ||  | ||  |||||||  |  |||  |   ||    ||
Db        794 GGGGGGGGTGCGGAGGGGGGGGGCGCCGGCGGGCGGCGGGCGGCGGGCGTGGC-GGGGCGG 852

Qy        492 CGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCCGGCACCGGGGT 551
               |   ||||||  |  | |||||  ||  | || |  || |||  | |.  || |||  ||
Db        853 GGGGGCGGGGGGGGGAGGGGGCGGGGGCGGCGGCGNGGGAGTAGCGGGGGCGCGCCGCGGG 912

Qy        552 GCGGGCCCGGCGGATCCCGG 571
              | ||  | | |||  |  ||
Db        913 GGGGCGCGGCCGGCGCGGGG 932


RESULT 14
BG441241/c
LOCUS       BG441241                925 bp    mRNA  linear   EST 15-MAR-2001
DEFINITION  GA__Ea0012G10f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Ea0012G10f, mRNA sequence.
ACCESSION   BG441241
VERSION     BG441241.1  GI:13350893
KEYWORDS    EST.
SOURCE      Gossypium arboreum.
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
            ,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence start: 39
            High quality sequence stop: 903.

FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Gossypium arboreum"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Ea0012G10f"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT       57 a    687 c    133 g     42 t      6 others
ORIGIN

  Query Match             8.0%;  Score 73.8;  DB 12;  Length 925;
  Best Local Similarity   45.9%;  Pred. No. 0.0006;
  Matches  252;  Conservative    0;  Mismatches  297;  Indels    0;  Gaps    0;

Qy         16 GGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCCTCTCCGGTGTTC 75
              ||   | | |||  | |||| |||   || | | |    |       | | ||| |  |
Db        773 GGGGCGTGGGGGGGGGGGGCGGGGGGCGGGGGGGGGGGGGGGGGGGGGGCGCGGGGGGGC 714

Qy         76 GCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGACTGGTCTGTGGTG 135
              |  ||  ||  |  |||| |    | ||    || |  | ||| ||   ||    | || |
Db        713 GGGCGTGGGGGGGGGGCGGGGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 654

Qy        136 GACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGATGTGGTCCAGCCT 195
              |     | | |  |||  | |||||    || | ||| | ||| |  | |  |     |
Db        653 GGGGGGCGGGGGGGGGGGGGGGGGGGCGGCGGGGGCGGGGGGGGGGGGGCGGGGGGGG 594

Qy        196 GTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTACGGTATCGAACCC 255
              | | |  | |    || |    ||  |    | || |        ||    |   ||
Db        593 GGGGGGCGCGGGCGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGCGGGGGGGGGCC 534

Qy        256 GACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTGTGTGGGGCGCTG 315
              | | ||     |||      |  ||| | |  |||||  |||   | |  |||| |  |
Db        533 GGGGGGGGGGGGGGCGGGGGCGGGGGGGGGGGCGCGGGGGCGGGGGGGGCGGGGGGGGG 474

Qy        316 AGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTGGCCGCTGTGCGG 375
              |   |  ||  | || |  |   | |  || ||| | | ||  ||  |  |   |  |
Db        473 GGGGGGGGGGGGGGGGGGGCGGGGGGGGGCGGGGGGGGCGGGGGGGGGGGGGGGGCGGC 414

Qy        376 GGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTGGAGCAGCTCATTGGT 435
              | |||| ||    ||     |  | |    |  |  || || | |     |   ||
Db        413 AGGCGGGCCGCCGGGGGGGCGCGGCGGGGGGGGGGGGGGGGGGGGCGGGGCGGCGGGGGC 354

Qy        436 GAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCCCGCTCCACCGCCGTC 495
              | | |||| |||||| ||  | ||| ||||  ||     ||     ||    | |  | |
Db        353 GGGGGGTGCGCGGGGGGGCCGCGGGGGGCGTGGGGAGCGGGATGGGGCGGGGCGGGGGGC 294

Qy        496 TCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCCGGCACCGGGGTGCGG 555
                | |||  || | |||||||  | || |  |||| | | || ||  ||||| | ||
Db        293 GGGAGGGGGGCGGGGGCGGTGGGGGGGCGGGGGCGGGCGGGGCGGGGCGCGGGGGGGGG 234

Qy        556 GCCCGGCGG 564
                || ||
Db        233 AGGGGGGGG 225


RESULT 15
AQ895329
LOCUS       AQ895329                918 bp    DNA     linear   GSS 10-NOV-1999
DEFINITION  HS_4832_A2_H01_T7A CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=4832 Col=2 Row=O, DNA sequence.
ACCESSION   AQ895329
VERSION     AQ895329.1  GI:6351435
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 918)
```

```
Db        882 GCGGGGGGGGGGGGGCGGGGGGGGGGGGGGGGGCGGGGGGGCGGCGCGGGCGGGGGGGCGGC 941

Qy        501 GGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCCGGCACCGGGGTGCGGG 556
Db        942 GCNTGGTGGGGGGGGGCGCGGGGCGGCGGCGGGGGCGCCCGCCGGGGGGGCGGGCG 997


RESULT 12
AG078502
LOCUS       AG078502               1137 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-073N07.R, genomic survey sequence.
ACCESSION   AG078502
VERSION     AG078502.1  GI:16630304
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-073N07.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1137)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1137
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-073N07.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT       23 a     13 c    768 g     24 t    309 others
ORIGIN

  Query Match             8.1%;  Score 74.6;  DB 17;  Length 1137;
  Best Local Similarity   36.5%;  Pred. No. 0.00043;
  Matches  203;  Conservative    0;  Mismatches  353;  Indels    0;  Gaps    0;

Qy         16 GGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCCTCTCCGGTGTTC 75
Db        330 GGNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGTTTTTAGTTTNNTTTTTNGGGGNGNG 389

Qy         76 GCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCCGTGGGTGGACTGGTCTGTGGTG 135
Db        390 GNGGGGGGGGGGNGNNGGGGGGGGGNNGNNGGGGNGGGGGGGGGGGGNGGNGGGGNNGNG 449

Qy        136 GACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGATGTGGTCCAGCCT 195
Db        450 NNNGGNNGNNNNGGGGNGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGNGGGGNNNNNNG 509

Qy        196 GTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTACGGTATCGAACCC 255
Db        510 GGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGNGGGGNNGNNGGGGNGNNGGGGGG 569

Qy        256 GACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTGTGTGGGGCGCTG 315
Db        570 NNGGNGGGGNNGGGNNGGGGNNGGGGGGGGGGGGGNNNGGGGGGGGNGGGGGNGGGNNG 629

Qy        316 AGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTGGCCGCTGTGCGG 375
Db        630 GNNNNNNNGGNGNNGGGGGNGGGNGNGGGGGNGGGGGGGNGGGGGGGGGNGGGGNNNN 689

Qy        376 GGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTGGAGCAGCTCATTGGT 435
Db        690 NGNGGGGGNGGGGGGNNGGGGGGGGGGGGGGGGGNNNNNGGGGNNNGGGGGGGGGNGGGG 749

Qy        436 GAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCCCGCTCCACCGCCGTC 495
Db        750 GNNNGGGGGGGGGGGGGNGNGGNGNGGGNNGGNNGGGGGNGGGGGGGNGNNGNNNGGNGGN 809

Qy        496 TCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCCGGCACCGGGGTGCGG 555
Db        810 GNNGGGGGGNNGGGGGGGGGNGNGGGGGGGNGGGNNGGNGNNNGGGGGGGGGGGGGGGGG 869

Qy        556 GCCCGGCGGATCCCGG 571
Db        870 GNGGGGGNGNGGNNGG 885


RESULT 13
AG137560
LOCUS       AG137560               1025 bp    DNA     linear   GSS 04-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-151H16.R, genomic survey sequence.
ACCESSION   AG137560
VERSION     AG137560.1  GI:16667238
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-151H16.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1025)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .1025
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-151H16.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT       28 a    163 c    775 g     25 t     34 others
ORIGIN

  Query Match             8.0%;  Score 74;  DB 17;  Length 1025;
```

```
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN12I14"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
BASE COUNT       83 a    566 c    179 g     93 t    180 others
ORIGIN

  Query Match             8.1%;  Score 75;  DB 17;  Length 1101;
  Best Local Similarity   42.2%;  Pred. No. 0.00036;
  Matches  237;  Conservative   25;  Mismatches  299;  Indels    0;  Gaps    0;

Qy         76 GCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGACTGGTCTGTGGTG 135
              ||  || ||  |  |||| :|| |  | |       || |:|  || |    ||   | || |
Db        608 GCGGGGGGGGGGNSGGCGNSCGGCGGGCGGGGCGCCCSGCCGGCGCGGCGGCGGGCGGCG 549

Qy        136 GACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGATGTGGTCCAGCCT 195
              | |  :  | | ||||  |:||| |  |  | |  ||   |||     | || |  :
Db        548 GGCGGSGGGGGGCGGGGGGSGGGCGGCGGCGGGCCGGGCGGGGGGCGGGGGGCCGCSGGG 489

Qy        196 GTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTACGGTATCGAACCC 255
                |  |    | |    || |       |  |  | | |          ||    |
Db        488 CCGGGGGGGGGGGGGGGGGSGGGGGGCGGGGGGCSGGGCGCSGGGCGGGGGGGGGCGGGG 429

Qy        256 GACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTGTGTGGGGCGCTG 315
              | |  || |    ||    ||: ||| | |   | ||  | :   | | | |||| || |
Db        428 GGCCGGGCCGGCCGCGGSGCCSGGGGGGGGGGGGGGGGGGGGSGGGGGGGGGGGGGGCSG 369

Qy        316 AGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTGGCCGCTGTGCGG 375
              |  |  |:  |:|||  |   |  |||:  :|  :: ||| |  ||  |  | | ||
Db        368 GGGGGGGGGSGGGSGGCGGSGGGCGGGGCGSGCSGGGSSGGGGGGGGGGGGGGGGGGGGG 309

Qy        376 GGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTGGAGCAGCTCATTGGT 435
              ||: |||| ||   ||      | |  |   |     :| :| || ||         |:
Db        308 GGSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGSGGSGGGGGGCSCGGGGGGGSG 249

Qy        436 GAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCCCGCTCCACCGCCGTC 495
              | | || ||| |||| ||  | |||| || ||:||      ||  :  |      | | :|
Db        248 GGGGGGGGGGGGGGGGGGGSGGGGGSGGGGGSGGSGGGGGGGGSGGGGGGGCCGGGSGGG 189

Qy        496 TCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCCGGCACCGGGGTGCGG 555
              | ||||  |  | || ||  |: || || |  |||:    | |  ||     |||: || |
Db        188 GCCGGGGGGGGGGGGGGGGSGCGGGGCGGGGGCSGGGCGCGGGGGGGGGGGGSGGCCG 129

Qy        556 GCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTGCAGCCCCTGCGGGAG 615
              |  || ||   || |  |      | |  || | | | | | || |  |  | |||  | |
Db        128 GGGGGGGGGCGGCCCGCGGCGGSCGGCCGGCGCGGGGCGCGCGGGCGGGGCNCGCGNGGC 69

Qy        616 GAGTTGCTGGAGCTGCTGGGG 636
              | |  |  || :  |  |:||
Db         68 GGGCGGGCGGSSGGGGSGSGG 48


RESULT 11
BQ670959
LOCUS       BQ670959                997 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8203742 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6255995
            5', mRNA sequence.
ACCESSION   BQ670959
VERSION     BQ670959.1  GI:21781793
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 997)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2407  row: p  column: 12
            High quality sequence start: 156
            High quality sequence stop: 281.
FEATURES             Location/Qualifiers
     source          1. .997
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6255995"
                     /clone_lib="NIH_MGC_102"
                     /tissue_type="epidermoid carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G).  Library constructed
                     by Ling Hong in the laboratory of Gerald M. Rubin
                     (University of  California, Berkeley) using ZAP-cDNA
                     synthesis kit (Stratagene) and Superscript II RT (Life
                     Technologies). Note: this is a NIH_MGC Library."
BASE COUNT       56 a    211 c    579 g     77 t     74 others
ORIGIN

  Query Match             8.1%;  Score 74.8;  DB 14;  Length 997;
  Best Local Similarity   45.7%;  Pred. No. 0.00039;
  Matches  245;  Conservative    0;  Mismatches  288;  Indels    3;  Gaps    1;

Qy         21 GGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCCTCTCCGGTGTTCGCCCG 80
              ||| |||  | |||  |||  ||| | | |  |  |  |                |  |  |.
Db        465 GGGNGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGNGGGNGGGGGGGGGGGGGGGGG 524

Qy         81 GCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGACTGGTCTGTGGTGGACAT 140
              | |   |  | | | |    ||     ||||  | ||| ||   ||   |  | ||
Db        525 GGGNGGGGGGNNGNGNGNNGGGGGGGGGCGGGGGGGGGGGGGGGGGGGGCGGGGGGGG 584

Qy        141 CCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGATGTGGTCCAGCCTGTGCT 200
              |  |  |||  | |||||  | || | | | | |||  |  | ||     |  | ||
Db        585 NGGGGGGGGGGGGGGGGGGGNGGGGGGGGNGGGGGGGGGGGNGGGGGGGGGGGGGGCG 644

Qy        201 GTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTACGGTATCGAACCCGACGC 260
              |    |  || |   ||| |   | || |        ||    |  | | ||
Db        645 GGGGGGGGGGGGGGNGGGGNGGGCGGGGGGGGGGGGGCGGNGGGGGGGGGGGCGGNCGG 704

Qy        261 GGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTGTGTGGGGCGCTGAGCCT 320
              ||    ||      |  ||| |  ||| |  | |   | | | | | | | | | |
Db        705 GG---GGGGGGCGGCGGGGGGGGGCGGCGCCGGGGGGGGGGGGGGCGCGGGGCGGGGGGG 761

Qy        321 GAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTGGCCGCTGTGCGGGGCCG 380
              |  ||  | || |  | | |  ||| | |      ||| || |||||  | | | | |
Db        762 GGGGGGGGGGGGGGGGCCCGGGGCGGCGGGGGCGGGGGGGCGGGCCGGCGGGGGGCGGGC 821

Qy        381 GGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTGGAGCAGCTCATTGGTGAGCG 440
              ||||||   ||     | || || |    ||  ||  | |  |  ||     || | | |
Db        822 GGGCGGGGGGGGGGNNGGGGCGGCCGGGCNCCNGGGGCGGGCNGGGGCGGGGCGGGGCGGG 881

Qy        441 GTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCCCGCTCCACCGCCGTCTCGGG 500
```

```
Qy        196 GTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTACGGTATCGAACCC 255
              | | |  | |   || |    |||  | |    |   | | |  ||   |
Db        413 GCGGGCGCGGGGGGGGGGGGGGGGNCGGCGNCGCGNNCCCCNCGCCCCCGGGNNCGGCGCN 472

Qy        256 GACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTGTGTGGGGCGCTG 315
                ||||  |  || |   | |  |     | || || | |||    | |     | | |
Db        473 NNGGCGGNGCCCGGGCGCGCGCCNGCGCCGGGCGGGGGCNCGCCCCCGCGGNNCNCCCGG 532

Qy        316 AGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTGGCCGCTGTGCGG 375
                |   | | || |  | |  | ||  | |||| |  || ||  |  |||
Db        533 CNAAGGCCGGGGGGGGGCCCGCGGGGGGCGCCGGGGGCGGGGGGCGGNCGGCGCCCGG 592

Qy        376 GGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTGGAGCAGCTCATTGGT 435
              |||  ||  | |||  || |     ||||  ||  |||| || |  |  | |  |
Db        593 GGCGGGCGGGGCGGGGGCCGCCGGCCGGCGCGGGCCCCGGCGCCGCGGNGGGGGGGGCGC 652

Qy        436 GAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCCCGCTCCACCGCCGTC 495
              |  ||| ||  ||||        | | |||||||      |||  | ||| |  |
Db        653 GGCCGGGGGNGGGGGNNCGGNCGCGGGGCGGCGCGGCGGGGCGGCGGCCCGGCGCGNNGG 712

Qy        496 TCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCCGGCACCGGGGTGCGG 555
                |||    || | ||||  ||  | ||     |||   | |  |||       |  | ||
Db        713 GGGGGCCCGGCGGCGGCGCGGGCGGGGGCCGNGGCCCCCNGCCCCGCGCGGCAGCGGGGG 772

Qy        556 GCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTGCAGC 604
                ||  |  |    ||     ||    | |    |||  |  ||||
Db        773 CCCGNGGNGCGGCGCCGCGGAGGGCCCCGCCGCGCGGCCCCCGCCCAGC 821
```

RESULT 9
BI888042

```
LOCUS       BI888042                811 bp    mRNA    linear   EST 12-OCT-2001
DEFINITION  ZF637-1-002297 Zebrafish shield stage whole embryo cDNA library
            MPMGp637 Danio rerio cDNA clone MPMGp637_10F13;MPMGp637F1310 5',
            mRNA sequence.
ACCESSION   BI888042
VERSION     BI888042.1  GI:16095313
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 811)
  AUTHORS   Clark,M., Aanstad,P., Hennig,S., Johnson,S.L. and Lehrach,H.
  TITLE     EST sequencing of a zebrafish shield stage cDNA library normalised
            by oligonucleotide fingerprinting
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Hennig S
            laboraty 123, dept.Lehrach
            Max-Planck-Institut fuer Molekulare Genetik
            Ihnestr.63-73, D-14195 Berlin, Germany
            Tel: +49 30 8413 1612
            Fax: +49 30 8413 1380
            Email: hennig@molgen.mpg.de
            5' EST sequencing of clones from a zebrafish shield stage library,
            normalised from 55,000 starting clones by oligonucleotide
            fingerprinting
            High quality sequence stop: 811.
FEATURES             Location/Qualifiers
     source          1. .811
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="MPMGp637_10F13;MPMGp637F1310"
                     /clone_lib="Zebrafish shield stage whole embryo cDNA
                     library MPMGp637"
                     /tissue_type="whole embryo"
                     /dev_stage="shield stage, 6 hrs post-fertilisation"
                     /lab_host="E.coli, XL1 blue MRF"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI;
                     oligo-dT-NotI primed, SalI adaptors, directionally cloned,
                     library normalised by oligonucleotide fingerprinting"
BASE COUNT       47 a    219 c    519 g     16 t     10 others
```

```
ORIGIN

  Query Match             8.2%;  Score 75.2;  DB 13;  Length 811;
  Best Local Similarity   46.4%;  Pred. No. 0.00033;
  Matches  327;  Conservative    0;  Mismatches  363;  Indels   14;  Gaps    2;

Qy         18 ACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCCTCTCCGGTGTTCGC 77
              || ||| |||  | |||| |||  ||| | | |   || ||   ||     || |
Db        119 ACGGGGGGGGGGGGGGGCGGGGGGGGGGGCGGGGGCGCGCGGGGCCCGGGTGGGGGGGCG 178

Qy         78 CCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGACTGGTCTGTGGTGGA 137
                || ||   | |||| | | || ||     || ||        | |  | ||| || ||
Db        179 GGGGGGGTGACGGGCGGGGGGGGGGGGGGGCGCCCCCCCCCCCCGCCGCGCCTGCGGGGGG 238

Qy        138 CATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGATGTGGTCCAGCCTGT 197
               ||  | |||   || |||||  | || | |||  || | ||    ||  ||
Db        239 GGCCCCCCCCCGCCGCGGGGGGGCGGGGGGGGGGGGCCCGCGGCGGGCCCCCCCTCCCCCC 298

Qy        198 GCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTACGGTATCGAACCCGA 257
              |    |   |    ||  | | | | |||  |        |    |     |
Db        299 CCCCCCCGCGGGGCGGGGCGGGGGGGGGGGGCGGGGGGGGGGGCGGCGGGGGGGGGGGG 358

Qy        258 CGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTGTGTGGGGCGCTGAG 317
              | |||     ||          |||  | |   | |||  | |   | | | || |||   |
Db        359 GGGGGGGGGGGGGGGGGGGGGGGGGGGGCGGGGGGGGGGTGGGGGGGGGGCGCGGGGGC 418

Qy        318 CCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTGGCCGCTGTGCGGGG 377
                |  ||  |||| |  |   |  | |  |  | |||  |  ||  |  | ||||||
Db        419 GGGGGGGGGGGCGGGGGGGCGGGGGGGGGCGGGGGGGGGGGGGGGGGGGGGGGGCGGGG 478

Qy        378 CCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTGGAGCAGCTCATTGGTGA 437
                |||  |||  ||      | |  |            || ||  |  || |      |  |
Db        479 GGCGGGGGGCGGGGGGGGGGCGGGG----------GGCGGGGTTGCTGTGGGGGGCGGG 527

Qy        438 GCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCCCGCTCCACCGCCGTCTC 497
              |||| | |||| ||||  |  || || |||||     |||||  | || |  | | ||
Db        528 GCGGCGTGCGGCGCGGGGGGCGGGGGGGGCGGGGGCCGGCCGCGGCGCAGCGGGCGGGGG 587

Qy        498 GGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCCGGCAC---CGGGGTGCG 554
              |||||  | || || |  ||  | || |  ||||   || ||  || |   ||||  |||
Db        588 GGGGGTGGGCGGGGGGCGGGCGGGGGCGGGGGCGCGGTGCGCGCGCGCGGGCGGGCGGCG 647

Qy        555 GGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTGCAGCCCCTGCGGGA 614
              |||  | | |  | | |  |  |||| ||| | ||        | || ||  |  || |
Db        648 GGCGGCGGGCGCCGCAGCGGGGAGTGCCGCGCGCGGCAGGGGCGGGCCGCGGCGCCGCGG 707

Qy        615 GGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGCGTGCCGTTCTTCTC 674
              |  |  || || || |  | |||    |   ||| | |   ||||    ||  |
Db        708 GCGGCCGCGGGCGCGGGCGCGGGGACGCCCGGCGCCGGCAGGCGGCCGCGCGCGCGCGGT 767

Qy        675 CACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCG 718
              |   |  |  || ||  | | ||| ||| |   || |   ||||
Db        768 CCGCGCCGGAGGGACGGGCCGGGAGACCGCCGGCCGGCCGGCCG 811
```

RESULT 10
CNS0153F/c

```
LOCUS       CNS0153F                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN12I14 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL104949
VERSION     AL104949.1  GI:5616963
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
```

```
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM1951  row: f  column: 06
          High quality sequence stop: 379.
FEATURES             Location/Qualifiers
     source          1. .1293
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5455109"
                     /clone_lib="NIH_MGC_98"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT       71 a    509 c    495 g     54 t    164 others
ORIGIN

  Query Match             8.5%;  Score 78;  DB 14;  Length 1293;
  Best Local Similarity   40.4%;  Pred. No. 0.0001;
  Matches  333;  Conservative    0;  Mismatches  482;  Indels   10;  Gaps    2;

Qy         91 GCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGACTGGTCTGTGGTGGACATCCTGCGCCGG 150
Db       1244 GCGGGCGNGGCGGGGGCGGGGCCGCGGCNCCGCCGGGCGGCGGGGCGGGGCGGGCGCCGG 1185

Qy        151 GACGCGGGGGATGCGGTGTGGGAGCGGGCCGATGTGGTCCAGCCTGTGCTGTTCAGCGTC 210
Db       1184 GCCGCGCGGCNGGGCCGCCGCGGGCCGCGCGGCGCGGCGGGCGCGCGGGCGCGGCNCGGC 1125

Qy        211 ATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTACGGTATCGAACCCGACGCGGTCCTTGGC 270
Db       1124 GNGGCGCGGNCGCGGGGCCGGGGGGCGCGGCGCNCGCCGCGGCGCGGCGCGGGCGCGGCG 1065

Qy        271 CATTCCCAGGGCGAGATCGCGGCCGCGCATGTGTGTGGGGCGCTGAGCCTGAAGGACGCG 330
Db       1064 CGCGCGGNGNNGGGGNGGCGGGGGGCGGGGGCNCGNGCGCCGCGGCGGCCGGCGGGCGCG 1005

Qy        331 GCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTGGCCGCTGTGCGGGGCCGGGGCGGCATG 390
Db       1004 GGG---------GGGGCGCGGCGCGGGGGGGGGGCGGCGGGCGGGGGGGGGCCGCGGGC 954

Qy        391 GCCTCAGTGCCGCTGCCTGCCCAGGAGGTGGAGCAGCTCATTGGTGAGCGGTGGGCGGGG 450
Db        953 GGGNCGGCGGGCCNCGGCGCGCGNNGCGCGGCGGCGCGGCGGGGCGGGCGCCCGGGGCGG 894

Qy        451 CGGTTGTGGGTGGCGGCGGTCAACGGCCCCCGCTCCACCGCCGTCTCGGGGGATGCCGAG 510
Db        893 GCGGGG-GGGCGGGGNGGCNCGGGGGCGGCCCCGGGCNGCGCGGGGCGGGGGGGGGCNCG 835

Qy        511 GCGGTGGACGAGGTGCTGGCGTACTGTGCCGGCACCGGGGTGCGGGCCCGGCGGATCCCG 570
Db        834 GGGGCGCCCGGGGGNCGCGCGGCCGGGGCCGCGGCGGGGGCCGCGGGGGCCGGGGGGCG 775

Qy        571 GTCGACTATGCCTCGCACTGCCCCCATGTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTG 630
Db        774 CGCCGCGCGCGGGGGGCGGGCGCGGGCGNCCCCCCCGCGGGGGGGGGCCGCCGGGCCGGC 715

Qy        631 CTGGGGGACATCAGCCCGCAGCCGTCCGGCGTGCCGTTCTTCTCCACGGTGGAGGGCACC 690
Db        714 GNGGGGCCGNGGCGGGCGGCGCCGCGCGGGGGGNGGCCCCGGNGGCGCCCCGCGGGGCCN 655

Qy        691 TGGCTGGACACCACAACCCTGGACGCCGCCTACTGGTACCGCAACCTGCACCAGCCGGTC 750
Db        654 CCGGNGGGCCCCCCCGCCGCGNNNGCCGCCCCNNCGGCCCCCCCCNNCCCCNNNCNNNN 595

Qy        751 CGTTTCAGCGATGCCGTCCAGGCCCTGGCGGATGACGGACACCGCGTCTTCGTCGAAGTC 810
```

```
Db        594 CCNNNCCCNNNCNNNCCCNCNCCCNNNNNNNNNCCCCCCCCNNNCCCCCCCNNCNNCNCC 535

Qy        811 AGCCCCCACCCCACCCTCGTCCCCGCCATCGAAGACACCACCGAAGACACCGCCGAAGAC 870
Db        534 CCCNCCCCNCCCCNNCCCCCNCCCCCCCCCCNCNNCCCCCCCNCCCGCCCCNNCCCCCCCC 475

Qy        871 GTCACCGCGATCGGCAGCCTCCGCCGCGGCGACAACGACACCCGC 915
Db        474 NNNNNNCCCCCCNNCCCCCCNCCCCNCNNNNNNCNNNNNNNCNC 430


RESULT 8
AG126210
LOCUS       AG126210                842 bp    DNA     linear   GSS 04-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-136L05.R, genomic survey sequence.
ACCESSION   AG126210
VERSION     AG126210.1  GI:16655375
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-136L05.R.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 842)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: M13Rev
             LIBRARY
                Vector    : pKS145
                R.Site 1  : SacI
                R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .842
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-136L05.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT       32 a    214 c    479 g     14 t    103 others
ORIGIN

  Query Match             8.2%;  Score 75.4;  DB 17;  Length 842;
  Best Local Similarity   43.0%;  Pred. No. 0.00031;
  Matches  253;  Conservative    0;  Mismatches  336;  Indels    0;  Gaps    0;

Qy         16 GGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCCTCTCCGGTGTTC 75
Db        233 GGGNGNGGGGGGGNGCGGNNGNGGGGCNGGGGGGGGGGGCGGNGCCCGGGNGGGGGGGGG 292

Qy         76 GCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGACTGGTCTGTGGTG 135
Db        293 CCNGGGGGGGGGGGGGCGGGGCGGGGGGGGGCGGGCGGGGGGGGGGGGGGCGGGGGGG 352

Qy        136 GACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGATGTGGTCCAGCCT 195
Db        353 GGGGGGGCCCCCCNGGGGGGGGGGGGGGGGGGGCCGGGGGGGCCGCGCGGGGGGCGCGG 412
```

```
Db        893 GGGGGGGGGGCGCGGGGGGGGGCGCCGGCCGGGGGCCCCGGCGNNGNNCGGGCGGCGGGG 834

Qy        557 CCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTGCAGCCCCTGCGGGAGG 616

Db        833 GCGGCGGGCGCGCCNGGGGCCNCGGCNCNCCCNCNCCCCCCCCCCCCCCNCCCNNCCC 774

Qy        617 AGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGCGTGCCGTTCTTCTCCA 676

Db        773 NCNCNCCCCNNCCCCCCCCCCNNCCCCCCCNCCNNNCCNCCCCCCCNCCNNNNCCCCCNC 714

Qy        677 CGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCCTACTGGTACCGCAACC 736

Db        713 CCNCCCCCCCCCCCCCCCCCCNNNCCNCCNCCCNNNCCCNCCNNNCCCCNNNCNNCCCNN 654

Qy        737 TGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCGGATGACGGACACCGCG 796

Db        653 CCNNNCCCCNNCNNNNCCCCCCCCCCCCCCCNNNNCCCCCCCCCCCCNNCNNNNCCCCC 594

Qy        797 TCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATCGAAGACACCACCGAAG 856

Db        593 NCCCCNCCNNNCCCCCCCCNCCCCNCNCCCCCCCCCCNCNCCNCCNNCNCNCCNNCCNCCC 534

Qy        857 ACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGCGACAACGACACCCGC 915

Db        533 CCCNNCCCNNNNCCNCNNNNCCCCCCCNNCCNCCCNNCCCCCCCCNCNNNNNCCCCCCC 475
```

RESULT 6
AG171171/c
LOCUS AG171171 691 bp DNA linear GSS 09-JAN-2002
DEFINITION Pan troglodytes DNA, clone: RP43-040G20.TJ, genomic survey sequence.
ACCESSION AG171171
VERSION AG171171.1 GI:16700849
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee Male BAC Library clone:RP43-040G20.TJ.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library RPCI-43
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 691)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library RPCI-43 This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS
Sequencing: TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1. .691
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-040G20.TJ"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT 0 a 454 c 168 g 0 t 69 others
ORIGIN

Query Match 8.5%; Score 78.2; DB 17; Length 691;
Best Local Similarity 43.9%; Pred. No. 9.4e-05;
Matches 274; Conservative 0; Mismatches 341; Indels 9; Gaps 2;

```
Qy         13 CCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCCTCTCCGGTG 72

Db        630 CNGGGGCGGGGGGGGGCGGCGGCCGCGCCCCGGNGCCCGCCCGCGCGCGGGGGCGCCGGGG 571

Qy         73 TTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGACTGGTCTGTG 132

Db        570 GGCGGGGCCGGGGCGGGGGCGCGGGGGGGGGGCGGGGGGGGGCGGGCNGCGCGGGGCGGG 511

Qy        133 GTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGATGTGGTCCAG 192

Db        510 GGGGGCGGCGGGGNCCCGGCGCGCGCCGGGGGNGGGGGGGCGGGGGCGGGGGGGGGGGC 451

Qy        193 CCTGTGCTGTTCAGCGTCATGGTGTCTTT------GGCTGCTCTGTGGCGTTCCTACGGT 246

Db        450 NNCGGGGGGGGCGGNGGGGCGGGGCCCNNCCGGCGGGGGGGGGGGGGGCCGGCCCCCCGC 391

Qy        247 ATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTGTGT 306

Db        390 CCCCCCCCCGGGCGGGGNNNCNGGGCGNCCGGCGGCGGGGCCCGGGGCGCGGGCGGGCGG 331

Qy        307 GGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTGGCC 366

Db        330 GCCCCGGGGGGGGCGGGGGGGCGGGGCGGGGGGGGGGCGGCGGGCGCGCGGCCGGGGGGG 271

Qy        367 GCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTGGAGCAG 426

Db        270 CGGGGGCGGGGCGGGGGGGGGGCGGCCCCGCGGGGCGGGGGCGCGCGGGGGCGGGGGGGG 211

Qy        427 CTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTC---AACGGCCCCCGC 483

Db        210 GGGGGCGGGGGGCGGGGGGGGGGGCGGGGGGGGGGGCCCGGGGGCGGGCGGGGGGGCCCC 151

Qy        484 TCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCCGGC 543

Db        150 NCNNNCCCCNNNCCGCGCCNGGGCGCGGCGCNGGCNGGGGGGNGGGGGGGGGGGGGGGG 91

Qy        544 ACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTGCAG 603

Db         90 GNGGGGGGGGGGGGGGGGGGGGGGGGGGNGGNNNNNNNNNNNNNNNNNNNNNNNNNG 31

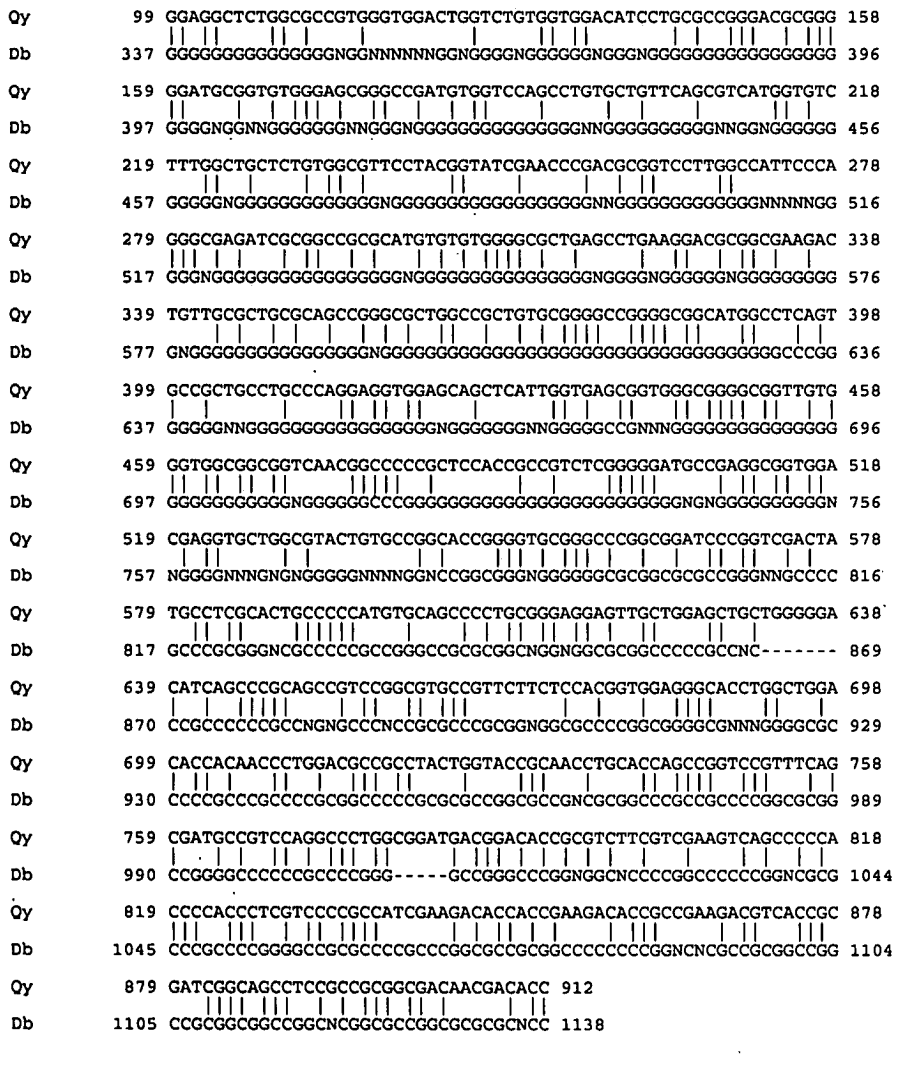
Qy        604 CCCCTGCGGGAGGAGTTGCTGGAG 627

Db         30 GNNNGNGGGGGGGGGNNGNGGGGG 7
```

RESULT 7
BM911120/c
LOCUS BM911120 1293 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6613421 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5455109 5', mRNA sequence.
ACCESSION BM911120
VERSION BM911120.1 GI:19361499
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1293)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be

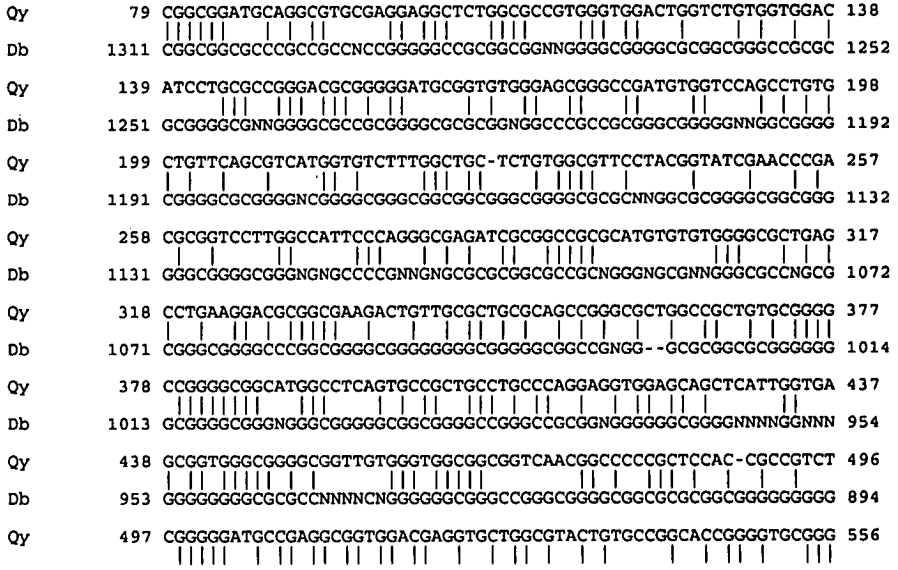
Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 40 a 356 c 673 g 27 t 268 others
ORIGIN

Query Match 8.8%; Score 80.6; DB 14; Length 1364;
Best Local Similarity 41.8%; Pred. No. 3.5e-05;
Matches 340; Conservative 0; Mismatches 462; Indels 12; Gaps 2;



RESULT 5
BM914111/c
LOCUS BM914111 1312 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6608135 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5478989 5', mRNA sequence.
ACCESSION BM914111
VERSION BM914111.1 GI:19364490
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1312)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1999 row: i column: 06
High quality sequence stop: 246.
FEATURES Location/Qualifiers
source 1. .1312
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5478989"
/clone_lib="NIH_MGC_98"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 46 a 381 c 531 g 41 t 313 others
ORIGIN

Query Match 8.6%; Score 79; DB 14; Length 1312;
Best Local Similarity 39.2%; Pred. No. 6.9e-05;
Matches 329; Conservative 0; Mismatches 506; Indels 4; Gaps 3;



```
Db        433 GGCGCGGGGGGGGGGGNGGCGGCGCGGGGGGGNNGCGGGGNGGCGCGGGGGGCGGGCGCGC 492

Qy        449 GGCGGTTGTGGGTGGCGGCGGTCAACGGCCCCCGCTCCACCGCCGTCTCGGGGGATGCCG 508
              ||  |  | ||| || ||||| |  ||| |   ||     ||  |    ||| |  |  |
Db        493 GGGCGGGGGGGGAGGGGGCGGGCGGCGGGCGGGGCGGGCGCGGGGGGGGGGCGGGGGGG 552

Qy        509 AGGCGGTGGACGAGGTGCTGGCGTACTGTGCCGGCACCGGGGTGCGGGCCCGGCGGATCC 568
              || |         || || ||    | |    ||  | || | | ||||  || ||
Db        553 GGGGGCCCCNGCGGGGGCGGGGNGGCGGGNNGGGGGCGGGCGCGGGGGCGGGGGGGGGG 612

Qy        569 CGGTCGACTATGCCTCGCACTGCCCCCATGTGCAGCCCCTGCGGGAGGAGTTGCTGGAGC 628
                |  |      ||  |||   | | ||  | ||  || | |      |     ||  | ||
Db        613 GCGGGGGGAGCGCGCCGCCGCGGCGCCGCGCGCGCCCGCGCCCCCCCGGCCCGCGCGCGC 672

Qy        629 TGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
                | || |  |  | || ||| ||    || |
Db        673 CGGCGGNGCGCGCCGGCGCGCGGCGCCGCGCC 704


RESULT 3
CNS0072Q
LOCUS       CNS0072Q                 932 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14B09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066742
VERSION     AL066742.1  GI:4945205
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .932
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR14B09"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
BASE COUNT      155 a    202 c    241 g     91 t    243 others
ORIGIN

  Query Match             8.9%;  Score 82.2;  DB 17;  Length 932;
  Best Local Similarity   32.5%;  Pred. No. 1.8e-05;
  Matches  161;  Conservative  113;  Mismatches  219;  Indels    2;  Gaps    1;

Qy        174 GCGGGCCGATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTG 233
              |||: |    | : :|   :|||| :::||   | : |:::|:|: :|:|::|:|  :|
Db        419 GCGSTCKCKCGCKCKCTGCSCTGTTTYKYTCKCTGCSCTSSBGYCKCKGKCKKCBCGCKG 478

Qy        234 GCGTTCCTACGGTATCGAACCCGACGCG--GTCCTTGGCCATTCCCAGGGCGAGATCGCG 291
              |  ::      ||  :     | |  | |  |     ||:| :|  | ||: | | :   :
Db        479 GTKBCKGGGKGGCBKGCKGGCGGGGGGGVCGRGGGSGGSCGSTGGCGGGKGGCGCKGCTS 538

Qy        292 GCCGCGCATGTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGC 351
              :    |:       | ||:::|: | ||::|   ||: :  ||:|  | ::::|:: ::| |
Db        539 KGGCTGBGSTGCGGGGGSKSGSGGSGCSBGCGGGGSTKGGSGGGGGSBSBTKYTBKGKGG 598

Qy        352 AGCCGGGCGCTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCC 411
              | ::|:| : |:| |  |   : ||: | |  ||:   ||: : | |: ||   | ||:
Db        599 CGGSSGSCCSGGSCRGGGGGCGSCGGSGGCGCGGGSSSCGCSCSCGCGSGGCGCGCCGCS 658

Qy        412 CAGGAGGTGGAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTC 471
              : :|  : || |  |: |   :   |:|   ::| :|   | | |  | : ||  |
Db        659 SGSGGCSSGGSGGSGSGCCGGCSSCGCCSGCCCCSSGCSCCCCCGCGCGSCGGSCCGCGC 718

Qy        472 AACGGCCCCCGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCG 531
              :|||::|  : || |: ||  | | :: |  :::| ||:|| :: ||  | |: ||:|
Db        719 SCSGGCSSCGCSCCCSCSCCCCSCCCCSSCGCSSSSGCGGSGGCSSGCGSCGCGSCGGSG 778

Qy        532 TACTGTGCCGGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGC 591
              : |   :  ::| |: ::|:|| ::::|:    :| :: |   |:: | :   |
Db        779 CGSCGGCGGSCGSSSGCGSGSSGSGCGSSSSGSSCGGGSGCSSGCSSGGSSGCCSSGSGG 838

Qy        592 CCCCATGTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAG 651
              ||||  | |  || :| :|| | ||    :| |  :| || | ||  :  : | | |: :
Db        839 CCCCSSGCGGCGCGSCSSCGCGSGGCCSGSCCGCCSCGGCCGCGGCGSGSSCGGCGGSCS 898

Qy        652 CCGTCCGGCGTGCCG 666
              |:  ||:|  |:::
Db        899 GCSSGCGSCCCGSSS 913


RESULT 4
BM810045
LOCUS       BM810045                 1364 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  AGENCOURT_6579545 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5469986
            5', mRNA sequence.
ACCESSION   BM810045
VERSION     BM810045.1  GI:19126868
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1364)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1976  row: b  column: 03
            High quality sequence stop: 178.
FEATURES             Location/Qualifiers
     source          1. .1364
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5469986"
                     /clone_lib="NIH_MGC_41"
                     /tissue_type="amelanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5469986"
/clone_lib="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 40 a 356 c 673 g 27 t 268 others
ORIGIN

Query Match 9.8%; Score 90.4; DB 14; Length 1364;
Best Local Similarity 44.9%; Pred. No. 5.9e-07;
Matches 301; Conservative 0; Mismatches 362; Indels 7; Gaps 2;

```
Qy        249 CGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTGTGTGG 308
Db       1255 CGNCCNCGNCGCGCCCCNGCGGGGGGCCNCGGGCCCGGCGGCGGNGGCGCCGCCCCGCNG 1196

Qy        309 GGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTGGCCGC 368
Db       1195 GGCG-GGCGGCCGGCGCCCCCGGCGGGCGNGCCGGGGGGCGCGGGCGGGGGGGGNNGCGG 1137

Qy        369 TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTGGAGCAGCT 428
Db       1136 NGCGCGCGCCGGCGCCGNGCCGGCCGCCGCGGCCGGCCGCGGCGNGNCCGGGGGGGGCCG 1077

Qy        429 CATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCCCGCTCCAC 488
Db       1076 CGGCGCCGGGCGGGGCGCGGCCCCGGGGCGGGCGCGNCCGGGGGGCCGGGGNGCCNCCGG 1017

Qy        489 CGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCCGGCACCGG 548
Db       1016 GCCCGGCCCCGGGGCGGGGGGCCCCGGCCGCGCCGGGGCGGCGGGCCGCGNCGGCGCCGG 957

Qy        549 GGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTGCAGCCCCT 608
Db        956 CGCGCGGGGGCCGCGGGGCGGGCGGGGGCGCCCCNNNCGCCCCGCCGGGGCGCCNCCGCG 897

Qy        609 GCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGCGTGCCGTT 668
Db        896 ------GGCGCGGNGGGCNCNGGCGGGGGGCGGGNGGCGGGGGCCGCGCCNCCNGCCGCG 843

Qy        669 CTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCCTACTGGTA 728
Db        842 CGGCCCGGCGGGGGCGNCCCGCGGGCGGGGCNNCCCGGCGCGCCGCGCCCCCNCCCGCC 783

Qy        729 CCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCGGATGACGG 788
Db        782 GGNCCNNNNCCCCCNCNCNNNCCCCNNCCCCCCCCCCNCNCCCCCCCCCCCCCCCCCCC 723

Qy        789 ACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATCGAAGACAC 848
Db        722 CCCCCGGGCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCNNNCGGCCCCCNNCCCCC 663

Qy        849 CACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGCGACAACGA 908
Db        662 CNCCCCCCCCCCCCCCCCCNNCCCCCCGGGCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 603

Qy        909 CACCCGCCGC 918
Db        602 CCCCCCCNCC 593
```

RESULT 2
AG036221
LOCUS AG036221 870 bp DNA linear GSS 01-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-012B17.F, genomic survey sequence.
ACCESSION AG036221
VERSION AG036221.1 GI:16563094
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC Library clone:PTB-012B17.F.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 870)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES Location/Qualifiers
source 1. .870
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-012B17.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 14 a 241 c 512 g 31 t 72 others
ORIGIN

Query Match 9.1%; Score 83.4; DB 17; Length 870;
Best Local Similarity 43.5%; Pred. No. 1.1e-05;
Matches 275; Conservative 0; Mismatches 356; Indels 1; Gaps 1;

```
Qy         30 GTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCCTCTCCGGTGTTCGCCCGGCGGATGCA 89
Db         73 GNNCGNCGGGCGGGGGGGGNTTTGTCGNGNNNCNNGCNGGGNNNCNNCNGGGNGGGNGNN 132

Qy         90 GGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGACTGGTCTGTGGTGGACATCCTGCGCCG 149
Db        133 GGGGCGGGGGGGGNNGGGGGGGGGGGCGCNNGGNGNGGGGGGGGCGNNTGTTTTGGGGGG 192

Qy        150 GGACGCGGGGGATGCGGTGTGGGAGCGGGCCGATGTGGTCCAGCCTGTGCTGTTCAGCGT 209
Db        193 GGGGTTGTGGGNNNNGGGGGGGGGGCGGGGGGGAGGGAGTTGTCCGTGGGTTTTTNGNGG 252

Qy        210 CATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTACGGT-ATCGAACCCGACGCGGTCCTTG 268
Db        253 NGGGGGNGNGGGGCGGGCGCGGGGGGGGNGGGGGGGGGGCGGGGGGGCGGGGCGGCCCGCG 312

Qy        269 GCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTGTGTGGGGCGCTGAGCCTGAAGGACG 328
Db        313 NCCCNCGNGGGGCGCCGGCGNNNGGCGNGCGCGGGGGGGGGGGGCGGGCGCGGGGGGGCG 372

Qy        329 CGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTGGCCGCTGTGCGGGGCCGGGGCGGCA 388
Db        373 CGCGCGCGCGGCGCGCGGGGCGGGGCGGGGGGGGCGCGGCCGGGCGGGGGGCGGGCGGCG 432

Qy        389 TGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTGGAGCAGCTCATTGGTGAGCGGTGGGCGG 448
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 1180.63 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_3148_4068
Perfect score: 921
Sequence: 1 gtcttcgtgtttcccggaca.....acaacgacacccgccgcttc 921

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	90.4	9.8	1364	14	BM810045	BM810045 AGENCOURT
	2	83.4	9.1	870	17	AG036221	AG036221 Pan trogl
	3	82.2	8.9	932	17	CNS0072Q	AL066742 Drosophil
	4	80.6	8.8	1364	14	BM810045	BM810045 AGENCOURT
c	5	79	8.6	1312	14	BM914111	BM914111 AGENCOURT
c	6	78.2	8.5	691	17	AG171171	AG171171 Pan trogl
c	7	78	8.5	1293	14	BM911120	BM911120 AGENCOURT
	8	75.4	8.2	842	17	AG126210	AG126210 Pan trogl
	9	75.2	8.2	811	13	BI888042	BI888042 ZF637-1-0
c	10	75	8.1	1101	17	CNS0153F	AL104949 Drosophil
	11	74.8	8.1	997	14	BQ670959	BQ670959 AGENCOURT
	12	74.6	8.1	1137	17	AG078502	AG078502 Pan trogl
	13	74	8.0	1025	17	AG137560	AG137560 Pan trogl
c	14	73.8	8.0	925	12	BG441241	BG441241 GA__Ea001
	15	73.6	8.0	918	17	AQ895329	AQ895329 HS_4832_A
	16	73.6	8.0	1213	13	BM560246	BM560246 AGENCOURT
	17	72.8	7.9	834	13	BI416657	BI416657 hasp001xm
c	18	72.8	7.9	1033	14	BQ674699	BQ674699 AGENCOURT
c	19	72.6	7.9	776	17	CNS010RY	AL099352 Drosophil
	20	72.6	7.9	1201	14	BQ063878	BQ063878 AGENCOURT
	21	72.6	7.9	1424	14	BQ720687	BQ720687 AGENCOURT
	22	72	7.8	1035	17	CNS017KO	AL108162 Drosophil
	23	72	7.8	1063	17	CNS040FA	AL268831 Tetraodon
	24	72	7.8	1171	14	BM914930	BM914930 AGENCOURT
c	25	71.8	7.8	1149	14	BM910848	BM910848 AGENCOURT
	26	71.6	7.8	827	17	AZ183795	AZ183795 SP_1002_A
	27	71.6	7.8	925	17	CNS0091P	AL053013 Drosophil
c	28	71.6	7.8	1161	17	AG032977	AG032977 Pan trogl
c	29	71.4	7.8	1052	14	BQ951911	BQ951911 AGENCOURT
	30	71.4	7.8	1136	14	BQ943816	BQ943816 AGENCOURT
	31	71.2	7.7	1247	14	BM809415	BM809415 AGENCOURT
c	32	70.8	7.7	1273	13	BM562099	BM562099 AGENCOURT
c	33	70.6	7.7	797	17	AG030595	AG030595 Pan trogl
c	34	70.6	7.7	888	12	BG809572	BG809572 mgct001xa
	35	70.4	7.6	1297	14	BQ064843	BQ064843 AGENCOURT
	36	70.2	7.6	798	13	BI416695	BI416695 hasp001xo
c	37	70.2	7.6	900	12	BG440857	BG440857 GA__Ea001
	38	70.2	7.6	918	12	BG809598	BG809598 mgct001xb
	39	70	7.6	739	17	AG080613	AG080613 Pan trogl
	40	70	7.6	1203	17	CNS015Y4	AL106054 Drosophil
	41	70	7.6	1278	14	BM802769	BM802769 AGENCOURT
	42	69.8	7.6	741	17	AZ194927	AZ194927 SP_1029_A
	43	69.8	7.6	947	17	AG030657	AG030657 Pan trogl
c	44	69.8	7.6	1006	10	BE034822	BE034822 ML04H02 M
	45	69.8	7.6	1151	14	BQ951254	BQ951254 AGENCOURT

ALIGNMENTS

RESULT 1
BM810045/c
LOCUS BM810045 1364 bp mRNA linear EST 05-MAR-2002
DEFINITION AGENCOURT_6579545 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5469986 5', mRNA sequence.
ACCESSION BM810045
VERSION BM810045.1 GI:19126868
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1364)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1976 row: b column: 03
High quality sequence stop: 178.
FEATURES Location/Qualifiers
source 1. .1364

THIS PAGE BLANK (USPTO)

```
Qy       481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
Db      6679 ACCGCCACCGTGGTTTCGGGCGACCCGACCCAGATCCAAGAGCTCGCTCAGGCGTGTGAG 6738

Qy       541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
Db      6739 GCCGACGGGGTCCGCGCACGGATCATCCCCGTCGACTACGCCTCCCACAGCGCCCACGTC 6798

Qy       601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
Db      6799 GAGACCATCGAGAGCGAACTCGCCGACGTCCTGGCGGGGTTGTCCCCCCAGACACCCCAG 6858

Qy       661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
Db      6859 GTCCCCTTCTTCTCCACCCTCGAAGGCGCCTGGATCACCGAACCCGCCCTCGACGGCGGC 6918

Qy       721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGG-- 778
Db      6919 TACTGGTACCGCAACCTCCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAAACCCTGGCC 6978

Qy       779 -CGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
Db      6979 ACCGACGAAGGCTTCACCCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTCACCATGGCG 7038

Qy       838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897
Db      7039 CTGCCCGAGACCGTCACCGGACTCGGCACCCTCCGCCGTGACAACGGCGGACAGCACCGC 7098

Qy       898 GGCGACAAC 906
Db      7099 CTCACCACC 7107
```

Search completed: June 18, 2003, 05:08:21
Job time : 112.933 secs

```
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-32

  Query Match             33.8%;  Score 311;  DB 9;  Length 11220;
  Best Local Similarity   61.3%;  Pred. No. 1.9e-61;
  Matches  557;  Conservative    0;  Mismatches  340;  Indels   12;  Gaps    3;

Qy          4 TTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCC 63
              |||||||| || ||||||||   ||||||||||||||||||||      || |||| |
Db       6205 TTCGTGTTCCCGGGACAGGGCACGCAGTGGGCCGGGATGGGTGCCGAACTCCTCGACGTG 6264

Qy         64 TCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGAC 123
              ||   || ||||||    |    |||    | ||||||||  || || ||  |||| ||||||
Db       6265 TCGAAGGAGTTCGCGGCGGCCATGGCCGAGTGCGAGGCCGCGCTCGCTCCGTACGTGGAC 6324

Qy        124 TGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGAT 183
              |||||  ||| || | || | || | || | | || |    |   | ||||||||  ||||
Db       6325 TGGTCGCTGGAGGCCGTCGTCCGACAGGCCCCCGGCGCGCCCACGCTGGAGCGGGTCGAT 6384

Qy        184 GTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTAC 243
              || |||||||| |||  ||||  |||||||| ||  ||||      | ||||    | ||
Db       6385 GTCGTCCAGCCCGTGACGTTCGCCGTCATGGTCTCGCTGGCGAAGGTCTGGCAGCACCAC 6444

Qy        244 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 303
              ||  |    ||  | || ||| | |||||| || ||||||||||||||| ||||||  | ||
Db       6445 GGGGTGACCCCGCAAGCCGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 6504

Qy        304 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 363
                || || |||||||||  | ||||| ||         ||  | |||||||||   | |  |
Db       6505 GCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGACCCTGCGCAGCAAGTCCATC 6564

Qy        364 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              | |||     |   ||||| ||||||||||    ||  |  ||||   |    ||  ||
Db       6565 GGCGCCCACCTCGCGGGCCAGGGCGGCATGCTGTCCCTCGCGCTGAGCGAGGCGGCCGTT 6624

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              | | |||   | |  |       ||  |  || |||  || || || ||||||||  ||
Db       6625 GTGGAGCGACTGGCCG------GGTTCGACGGGCTGTCCGTCGCCGCCGTCAACGGGCCT 6678

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              | ||||||  || ||||| ||  |      | |  | ||| |       |  ||||
Db       6679 ACCGCCACCGTGGTTTCGGGCGACCCGACCCAGATCCAAGAGCTCGCTCAGGCGTGTGAG 6738

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              | |  |||||| || || |||  ||||| |||||||| ||||| ||| || |||| ||
Db       6739 GCCGACGGGGTCCGCGCACGGATCATCCCCGTCGACTACGCCTCCCACAGCGCCCACGTC 6798

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              || || |   |   ||  |   ||  | |||| ||   |   ||| ||| |  ||
Db       6799 GAGACCATCGAGAGCGAACTCGCCGACGTCCTGGCGGGGTTGTCCCCCCAGACACCCCAG 6858

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              || || ||||||||||||  | || ||| ||||| |    |      | |||| |||| || |
Db       6859 GTCCCCTTCTTCTCCACCCTCGAAGGCGCCTGGATCACCGAACCCGCCCTCGACGGCGGC 6918

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGG-- 778
              |||||||||||||||| | ||| |  ||  | |||  |   |||||| |  ||||||

Db       6919 TACTGGTACCGCAACCTCCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAAACCCTGGCC 6978

Qy        779 -CGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
               | || || ||   |  |  ||||||||| |||||| ||||||||  ||||  |   ||
Db       6979 ACCGACGAAGGCTTCACCCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTCACCATGGCG 7038

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897
              |    || |||  |   |    || |.    |  |   |  | |||| | |  | ||||
Db       7039 CTGCCCGAGACCGTCACCGGACTCGGCACCCTCCGCCGTGACAACGGCGGACAGCACCGC 7098

Qy        898 GGCGACAAC 906
                |  || |
Db       7099 CTCACCACC 7107


RESULT 15
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-32

  Query Match             33.8%;  Score 311;  DB 10;  Length 11220;
  Best Local Similarity   61.3%;  Pred. No. 1.9e-61;
  Matches  557;  Conservative    0;  Mismatches  340;  Indels   12;  Gaps    3;

Qy          4 TTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCC 63
              |||||||| || ||||||||   ||||||||||||||||||||      || |||| |
Db       6205 TTCGTGTTCCCGGGACAGGGCACGCAGTGGGCCGGGATGGGTGCCGAACTCCTCGACGTG 6264

Qy         64 TCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGAC 123
              ||   || ||||||    |    |||    | ||||||||  || || ||  |||| ||||||
Db       6265 TCGAAGGAGTTCGCGGCGGCCATGGCCGAGTGCGAGGCCGCGCTCGCTCCGTACGTGGAC 6324

Qy        124 TGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGAT 183
              |||||  ||| || | || | || | || | | || |    |   | ||||||||  ||||
Db       6325 TGGTCGCTGGAGGCCGTCGTCCGACAGGCCCCCGGCGCGCCCACGCTGGAGCGGGTCGAT 6384

Qy        184 GTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTAC 243
              || |||||||| |||  ||||  |||||||| ||  ||||      | ||||    | ||
Db       6385 GTCGTCCAGCCCGTGACGTTCGCCGTCATGGTCTCGCTGGCGAAGGTCTGGCAGCACCAC 6444

Qy        244 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 303
              ||  |    ||  | || ||| | |||||| || ||||||||||||||| ||||||  | ||
Db       6445 GGGGTGACCCCGCAAGCCGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 6504

Qy        304 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 363
                || || |||||||||  | ||||| ||         ||  | |||||||||   | |  |
Db       6505 GCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGACCCTGCGCAGCAAGTCCATC 6564

Qy        364 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              | |||     |   ||||| ||||||||||    ||  |  ||||   |    ||  ||
Db       6565 GGCGCCCACCTCGCGGGCCAGGGCGGCATGCTGTCCCTCGCGCTGAGCGAGGCGGCCGTT 6624

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              | | |||   | |  |       ||  |  || |||  || || || ||||||||  ||
Db       6625 GTGGAGCGACTGGCCG------GGTTCGACGGGCTGTCCGTCGCCGCCGTCAACGGGCCT 6678
```

```
Qy        244 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 303
              ||  |    ||  | || ||| | ||||| || ||||||||||||||| |||||| | ||
Db       6445 GGGGTGACCCCGCAAGCCGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 6504

Qy        304 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 363
                || || |||||||| | ||||| ||       ||  | |||||||||  | |  |
Db       6505 GCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGACCCTGCGCAGCAAGTCCATC 6564

Qy        364 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              | |||    |  ||||| ||||||||||    ||  |  |||||    |    ||  ||
Db       6565 GGCGCCCACCTCGCGGGCCAGGGCGGCATGCTGTCCCTCGCGCTGAGCGAGGCGGCCGTT 6624

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              | | |||   | |  |     ||   |  || |||  || || || |||||||| ||
Db       6625 GTGGAGCGACTGGCCG------GGTTCGACGGGCTGTCCGTCGCCGCCGTCAACGGGCCT 6678

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              | ||||||  || ||||| ||  |     | |  | ||| |      |   ||||
Db       6679 ACCGCCACCGTGGTTTCGGGCGACCCGACCCAGATCCAAGAGCTCGCTCAGGCGTGTGAG 6738

Qy        541 GGCACCGGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              | |  |||||| || || |||   ||||| |||||||| ||||| ||| || |||| ||
Db       6739 GCCGACGGGGTCCGCGCACGGATCATCCCCGTCGACTACGCCTCCCACAGCGCCCACGTC 6798

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              || || |   |   ||  |   ||  | |||| ||   |   ||| ||| |  ||
Db       6799 GAGACCATCGAGAGCGAACTCGCCGACGTCCTGGCGGGGTTGTCCCCCCAGACACCCCAG 6858

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              || || |||||||||||  | || ||| ||||| |    |     | |||| |||| || |
Db       6859 GTCCCCTTCTTCTCCACCCTCGAAGGCGCCTGGATCACCGAACCCGCCCTCGACGGCGGC 6918

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGG-- 778
              |||||||||||||||| | ||| |  ||  | |||  |   ||||||  |  ||||||
Db       6919 TACTGGTACCGCAACCTCCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAAACCCTGGCC 6978

Qy        779 -CGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
               | || || ||   |  |  ||||||||| ||||||  |||||||  ||||  |   ||
Db       6979 ACCGACGAAGGCTTCACCCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTCACCATGGCG 7038

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897
              |    || |||  |   |   || |     |  |   |  | |||| | |  | ||||
Db       7039 CTGCCCGAGACCGTCACCGGACTCGGCACCCTCCGCCGTGACAACGGCGGACAGCACCGC 7098

Qy        898 GGCGACAAC 906
              |  || |
Db       7099 CTCACCACC 7107
```

```
RESULT 13
US-09-988-384B-32
; Sequence 32, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-32
```

```
  Query Match           33.8%;  Score 311;  DB 9;  Length 11220;
  Best Local Similarity   61.3%;  Pred. No. 1.9e-61;
  Matches  557;  Conservative    0;  Mismatches  340;  Indels   12;  Gaps    3;

Qy          4 TTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCC 63
              |||||||| || |||||||||   ||||||||||||||||||||||      || |||| |
Db       6205 TTCGTGTTCCCGGGACAGGGCACGCAGTGGGCCGGGATGGGTGCCGAACTCCTCGACGTG 6264

Qy         64 TCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGAC 123
              ||   || ||||||    |   |||   | ||||||||  || || || ||||  ||||||
Db       6265 TCGAAGGAGTTCGCGGCGGCCATGGCCGAGTGCGAGGCCGCGCTCGCTCCGTACGTGGAC 6324

Qy        124 TGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGAT 183
              |||||  ||| || | || | || | || | || |    |   | ||||||| ||||
Db       6325 TGGTCGCTGGAGGCCGTCGTCCGACAGGCCCCCGGCGCGCCCACGCTGGAGCGGGTCGAT 6384

Qy        184 GTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTAC 243
              || |||||||| |||  |||| |||||||| ||  ||||     | ||||    | ||
Db       6385 GTCGTCCAGCCCGTGACGTTCGCCGTCATGGTCTCGCTGGCGAAGGTCTGGCAGCACCAC 6444

Qy        244 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 303
              ||  |    ||  | || ||| | ||||| || ||||||||||||||| |||||| | ||
Db       6445 GGGGTGACCCCGCAAGCCGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 6504

Qy        304 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 363
                || || |||||||| | ||||| ||       ||  | |||||||||  | |  |
Db       6505 GCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGACCCTGCGCAGCAAGTCCATC 6564

Qy        364 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              | |||    |  ||||| ||||||||||    ||  |  |||||    |    ||  ||
Db       6565 GGCGCCCACCTCGCGGGCCAGGGCGGCATGCTGTCCCTCGCGCTGAGCGAGGCGGCCGTT 6624

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              | | |||   | |  |     ||   |  || |||  || || || |||||||| ||
Db       6625 GTGGAGCGACTGGCCG------GGTTCGACGGGCTGTCCGTCGCCGCCGTCAACGGGCCT 6678

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              | ||||||  || ||||| ||  |     | |  | ||| |      |   ||||
Db       6679 ACCGCCACCGTGGTTTCGGGCGACCCGACCCAGATCCAAGAGCTCGCTCAGGCGTGTGAG 6738

Qy        541 GGCACCGGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              | |  |||||| || || |||   ||||| |||||||| ||||| ||| || |||| ||
Db       6739 GCCGACGGGGTCCGCGCACGGATCATCCCCGTCGACTACGCCTCCCACAGCGCCCACGTC 6798

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              || || |   |   ||  |   ||  | |||| ||   |   ||| ||| |  ||
Db       6799 GAGACCATCGAGAGCGAACTCGCCGACGTCCTGGCGGGGTTGTCCCCCCAGACACCCCAG 6858

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              || || |||||||||||  | || ||| ||||| |    |     | |||| |||| || |
Db       6859 GTCCCCTTCTTCTCCACCCTCGAAGGCGCCTGGATCACCGAACCCGCCCTCGACGGCGGC 6918

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGG-- 778
              |||||||||||||||| | ||| |  ||  | |||  |   ||||||  |  ||||||
Db       6919 TACTGGTACCGCAACCTCCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAAACCCTGGCC 6978

Qy        779 -CGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
               | || || ||   |  |  ||||||||| ||||||  |||||||  ||||  |   ||
Db       6979 ACCGACGAAGGCTTCACCCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTCACCATGGCG 7038

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897
              |    || |||  |   |   || |     |  |   |  | |||| | |  | ||||
Db       7039 CTGCCCGAGACCGTCACCGGACTCGGCACCCTCCGCCGTGACAACGGCGGACAGCACCGC 7098

Qy        898 GGCGACAAC 906
              |  || |
Db       7099 CTCACCACC 7107
```

```
RESULT 14
US-09-836-821-32
; Sequence 32, Application US/09836821
; Publication No. US20030087405A1
```

```
Db       5645 CTCCCCGGGACC 5656


RESULT 11
US-09-793-708-19
; Sequence 19, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-19

  Query Match           34.5%;  Score 318;  DB 9;  Length 38506;
  Best Local Similarity   63.4%;  Pred. No. 4.3e-63;
  Matches  540;  Conservative    0;  Mismatches  300;  Indels   12;  Gaps    3;

Qy          4 TTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCC 63
              |||||||| ||||| |||||   |||||||||||| |||||||     |||||  | |||
Db       1783 TTCGTGTTCCCCGGGCAGGGCACGCAGTGGGCCGGCATGGGTGCCGAACTGCTGGACTCT 1842

Qy         64 TCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGAC 123
              ||  ||||||||||    |   |||   |  |||||||  || ||  | ||||  || |||
Db       1843 TCCGCGGTGTTCGCGGCGGCCATGGCCGAATGCGAGGCCGCACTCTCCCCGTACGTCGAC 1902

Qy        124 TGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGAT 183
              |||||  ||| || | || | || | || | | || |   |   | ||||||||| ||||
Db       1903 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 1962

Qy        184 GTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTAC 243
              || || |||||||||  ||||  |||||||| ||  |||||    ||||||    | ||
Db       1963 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 2022

Qy        244 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 303
              ||   |    ||| | |||||| | ||||| || ||||||||||||||| ||||||  | ||
Db       2023 GGGGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 2082

Qy        304 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 363
                 || || |||||||| | ||||| ||      ||  | |||||||||  | |  |
Db       2083 GCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGACCCTGCGCAGCAAGTCCATC 2142

Qy        364 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              |||||    |   |||  ||||||||||    ||  |  |||||   |   | |  ||
Db       2143 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAGCGAGGACGCCGTC 2202

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              | |||   | |  |       ||    |  || |||  || || ||  || ||||| |||
Db       2203 CTGGAGCGACTGGCCG------GGTTCGACGGGCTGTCCGTCGCCGCTGTGAACGGGCCC 2256

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              | ||||||  |||||  || ||  |||    | | || ||| |      |   ||||
Db       2257 ACCGCCACCGTGGTCTCCGGTGACCCCGTACAGATCGAAGAGCTTGCTCGGGCGTGTGAG 2316

Qy        541 GGCACCGGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              | |   ||||| || || |||   || || |||||||| || || ||| |||  || ||
Db       2317 GCCGATGGGGTCCGTGCGCGGGTCATTCCCGTCGACTACGCGTCCCACAGCCGGCAGGTC 2376

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              ||  | |   |   ||| |   ||| | || |  |   ||||||||||| |   |  ||
Db       2377 GAGATCATCGAGAGCGAGCTCGCCGAGGTCCTCGCCGGGCTCAGCCCGCAGGCTCCGCGC 2436

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              |||||||||||||| ||  | || ||| ||||| |   |     |    || |||| || |
Db       2437 GTGCCGTTCTTCTCGACACTCGAAGGCGCCTGGATCACCGAGCCCGTGCTCGACGGCGGC 2496

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGG-- 778
              |||||||||||||||||| ||| |  ||  | |||  |   ||||||| || ||||||
Db       2497 TACTGGTACCGCAACCTGCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAGACCCTGGCC 2556

Qy        779 -CGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
              | || ||. ||   |  |  |||||||||| |||||| ||||||| ||||  |   |||
Db       2557 ACCGACGAGGGCTTCACCCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTCACCATGGCC 2616

Qy        838 ATCGAAGACACC 849
              ||   |  |||
Db       2617 CTCCCCGGGACC 2628


RESULT 12
US-09-860-846-32
; Sequence 32, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-32

  Query Match           33.8%;  Score 311;  DB 9;  Length 11220;
  Best Local Similarity   61.3%;  Pred. No. 1.9e-61;
  Matches  557;  Conservative    0;  Mismatches  340;  Indels   12;  Gaps    3;

Qy          4 TTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCC 63
              |||||||| || |||||||||   |||||||||||||||||||||      || |||| |
Db       6205 TTCGTGTTCCCGGGACAGGGCACGCAGTGGGCCGGGATGGGTGCCGAACTCCTCGACGTG 6264

Qy         64 TCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGAC 123
              ||   || ||||||    |   |||   | ||||||||  || || ||  |||| ||||||
Db       6265 TCGAAGGAGTTCGCGGCGGCCATGGCCGAGTGCGAGGCCGCGCTCGCTCCGTACGTGGAC 6324

Qy        124 TGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGAT 183
              |||||  ||| || | || | || | || | | || |   |   | |||||||| ||||
Db       6325 TGGTCGCTGGAGGCCGTCGTCCGACAGGCCCCCGGCGCGCCCACGCTGGAGCGGGTCGAT 6384

Qy        184 GTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTAC 243
              || |||||||| |||  ||||  |||||||| ||  ||||     | ||||    | ||
Db       6385 GTCGTCCAGCCCGTGACGTTCGCCGTCATGGTCTCGCTGGCGAAGGTCTGGCAGCACCAC 6444
```

```
Qy          4 TTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCC 63
              |||||||| ||||| |||||   |||||||||||| |||||||      ||||| | |||
Db       3641 TTCGTGTTCCCCGGGCAGGGCACGCAGTGGGCCGGCATGGGTGCCGAACTGCTGGACTCT 3700

Qy         64 TCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGAC 123
              ||  ||||||||||   |   |||   |||||||  || ||  | ||||  || |||
Db       3701 TCCGCGGTGTTCGCGGCGGCCATGGCCGAATGCGAGGCCGCACTCTCCCCGTACGTCGAC 3760

Qy        124 TGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGAT 183
              |||||  ||| || | || | || | || | | || |   |   | ||||||| ||||
Db       3761 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 3820

Qy        184 GTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTAC 243
              || || |||||||||  ||||  |||||||| ||  |||||     ||||||    | ||
Db       3821 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 3880

Qy        244 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 303
              ||  |   ||| | |||||| | ||||| || ||||||||||||||| |||||| | ||
Db       3881 GGGGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 3940

Qy        304 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 363
                || || ||||||||| | |||| ||          ||  | |||||||||  | |  |
Db       3941 GCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGACCCTGCGCAGCAAGTCCATC 4000

Qy        364 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              |||||    |   |||  ||||||||||   ||  |  |||||    |  | |  ||
Db       4001 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAGCGAGGACGCCGTC 4060

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              | |||   | |  |       ||   | || ||| || || || || || ||||| |||
Db       4061 CTGGAGCGACTGGCCG------GGTTCGACGGGCTGTCCGTCGCCGCTGTGAACGGGCCC 4114

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              | ||||||  ||||| || || |||   || || |||||||| || || ||| |||  || ||
Db       4115 ACCGCCACCGTGGTCTCCGGTGACCCCGTACAGATCGAAGAGCTTGCTCGGGCGTGTGAG 4174

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              | |   ||||| || || |||   || || |||||||| || || ||| |||  || ||
Db       4175 GCCGATGGGGTCCGTGCGCGGGTCATTCCCGTCGACTACGCGTCCCACAGCCGGCAGGTC 4234

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              ||  | |   |  ||| |   ||| | || |  |  ||||||||||| |  |  ||
Db       4235 GAGATCATCGAGAGCGAGCTCGCCGAGGTCCTCGCCGGGCTCAGCCCGCAGGCTCCGCGC 4294

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              |||||||||||||| ||  | || ||| ||||| |  |   |   || |||| || |
Db       4295 GTGCCGTTCTTCTCGACACTCGAAGGCGCCTGGATCACCGAGCCCGTGCTCGACGGCGGC 4354

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGG-- 778
              |||||||||||||||||| ||| |  ||  | |||  |  ||||||  || ||||||
Db       4355 TACTGGTACCGCAACCTGCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAGACCCTGGCC 4414

Qy        779 -CGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
               | || || ||   |  | |||||||| |||||| ||||||||  ||||  |   |||
Db       4415 ACCGACGAGGGCTTCACCCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTCACCATGGCC 4474

Qy        838 ATCGAAGACACC 849
              ||   |  |||
Db       4475 CTCCCCGGGACC 4486


RESULT 10
US-09-988-384B-5
; Sequence 5, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 5
;   LENGTH: 37948
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-5

  Query Match            34.5%;  Score 318;  DB 9;  Length 37948;
  Best Local Similarity  63.4%;  Pred. No. 4.3e-63;
  Matches  540;  Conservative    0;  Mismatches  300;  Indels   12;  Gaps    3;

Qy          4 TTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCC 63
              |||||||| ||||| |||||   |||||||||||| |||||||      ||||| | |||
Db       4811 TTCGTGTTCCCCGGGCAGGGCACGCAGTGGGCCGGCATGGGTGCCGAACTGCTGGACTCT 4870

Qy         64 TCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGAC 123
              ||  ||||||||||   |   |||   |||||||  || ||  | ||||  || |||
Db       4871 TCCGCGGTGTTCGCGGCGGCCATGGCCGAATGCGAGGCCGCACTCTCCCCGTACGTCGAC 4930

Qy        124 TGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGAT 183
              |||||  ||| || | || | || | || | | || |   |   | ||||||| ||||
Db       4931 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 4990

Qy        184 GTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTAC 243
              || || |||||||||  ||||  |||||||| ||  |||||     ||||||    | ||
Db       4991 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 5050

Qy        244 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 303
              ||  |   ||| | |||||| | ||||| || ||||||||||||||| |||||| | ||
Db       5051 GGGGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 5110

Qy        304 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 363
                || || |||||||||  | ||||| ||          ||  | |||||||||  | |  |
Db       5111 GCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGACCCTGCGCAGCAAGTCCATC 5170

Qy        364 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              |||||    |   |||  ||||||||||   ||  |  |||||    |  | |  ||
Db       5171 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAGCGAGGACGCCGTC 5230

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              | |||   | |  |       ||   | || ||| || || || || || ||||| |||
Db       5231 CTGGAGCGACTGGCCG------GGTTCGACGGGCTGTCCGTCGCCGCTGTGAACGGGCCC 5284

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              | ||||||  ||||| || || |||   || || |||||||| || || ||| |||  || ||
Db       5285 ACCGCCACCGTGGTCTCCGGTGACCCCGTACAGATCGAAGAGCTTGCTCGGGCGTGTGAG 5344

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              | |   ||||| || || |||   || || |||||||| || || ||| |||  || ||
Db       5345 GCCGATGGGGTCCGTGCGCGGGTCATTCCCGTCGACTACGCGTCCCACAGCCGGCAGGTC 5404

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              ||  | |   |  ||| |   ||| | || |  |  ||||||||||| |  |  ||
Db       5405 GAGATCATCGAGAGCGAGCTCGCCGAGGTCCTCGCCGGGCTCAGCCCGCAGGCTCCGCGC 5464

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              |||||||||||||| ||  | || ||| ||||| |  |   |   || |||| || |
Db       5465 GTGCCGTTCTTCTCGACACTCGAAGGCGCCTGGATCACCGAGCCCGTGCTCGACGGCGGC 5524

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGG-- 778
              |||||||||||||||||| ||| |  ||  | |||  |  ||||||  || ||||||
Db       5525 TACTGGTACCGCAACCTGCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAGACCCTGGCC 5584

Qy        779 -CGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
               | || || ||   |  | |||||||| |||||| ||||||||  ||||  |   |||
Db       5585 ACCGACGAGGGCTTCACCCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTCACCATGGCC 5644

Qy        838 ATCGAAGACACC 849
```

```
Db       4061 CTGGAGCGACTGGCCG------GGTTCGACGGGCTGTCCGTCGCCGCTGTGAACGGGCCC 4114

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540

Db       4115 ACCGCCACCGTGGTCTCCGGTGACCCCGTACAGATCGAAGAGCTTGCTCGGGCGTGTGAG 4174

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600

Db       4175 GCCGATGGGGTCCGTGCGCGGGTCATTCCCGTCGACTACGCGTCCCACAGCCGGCAGGTC 4234

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660

Db       4235 GAGATCATCGAGAGCGAGCTCGCCGAGGTCCTCGCCGGGCTCAGCCCGCAGGCTCCGCGC 4294

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720

Db       4295 GTGCCGTTCTTCTCGACACTCGAAGGCGCCTGGATCACCGAGCCCGTGCTCGACGGCGGC 4354

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGG-- 778

Db       4355 TACTGGTACCGCAACCTGCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAGACCCTGGCC 4414

Qy        779 -CGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837

Db       4415 ACCGACGAGGGCTTCACCCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTCACCATGGCC 4474

Qy        838 ATCGAAGACACC 849

Db       4475 CTCCCCGGGACC 4486


RESULT 8
US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-5

  Query Match            34.5%;  Score 318;  DB 9;  Length 36778;
  Best Local Similarity   63.4%;  Pred. No. 4.4e-63;
  Matches  540;  Conservative    0;  Mismatches  300;  Indels   12;  Gaps    3;

Qy          4 TTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCC 63

Db       3641 TTCGTGTTCCCCGGGCAGGGCACGCAGTGGGCCGGCATGGGTGCCGAACTGCTGGACTCT 3700

Qy         64 TCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGAC 123

Db       3701 TCCGCGGTGTTCGCGGCGGCCATGGCCGAATGCGAGGCCGCACTCTCCCCGTACGTCGAC 3760

Qy        124 TGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGAT 183

Db       3761 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 3820

Qy        184 GTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTAC 243

Db       3821 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 3880

Qy        244 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 303

Db       3881 GGGGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 3940

Qy        304 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 363

Db       3941 GCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGACCCTGCGCAGCAAGTCCATC 4000

Qy        364 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420

Db       4001 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAGCGAGGACGCCGTC 4060

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480

Db       4061 CTGGAGCGACTGGCCG------GGTTCGACGGGCTGTCCGTCGCCGCTGTGAACGGGCCC 4114

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540

Db       4115 ACCGCCACCGTGGTCTCCGGTGACCCCGTACAGATCGAAGAGCTTGCTCGGGCGTGTGAG 4174

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600

Db       4175 GCCGATGGGGTCCGTGCGCGGGTCATTCCCGTCGACTACGCGTCCCACAGCCGGCAGGTC 4234

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660

Db       4235 GAGATCATCGAGAGCGAGCTCGCCGAGGTCCTCGCCGGGCTCAGCCCGCAGGCTCCGCGC 4294

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720

Db       4295 GTGCCGTTCTTCTCGACACTCGAAGGCGCCTGGATCACCGAGCCCGTGCTCGACGGCGGC 4354

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGG-- 778

Db       4355 TACTGGTACCGCAACCTGCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAGACCCTGGCC 4414

Qy        779 -CGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837

Db       4415 ACCGACGAGGGCTTCACCCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTCACCATGGCC 4474

Qy        838 ATCGAAGACACC 849

Db       4475 CTCCCCGGGACC 4486


RESULT 9
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-5

  Query Match            34.5%;  Score 318;  DB 10;  Length 36778;
  Best Local Similarity   63.4%;  Pred. No. 4.4e-63;
  Matches  540;  Conservative    0;  Mismatches  300;  Indels   12;  Gaps    3;
```

```
RESULT 6
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-30

  Query Match             34.5%;  Score 318;  DB 10;  Length 13842;
  Best Local Similarity   63.4%;  Pred. No. 4.9e-63;
  Matches  540;  Conservative    0;  Mismatches  300;  Indels   12;  Gaps    3;

Qy          4 TTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCC 63
              |||||||| |||||  |||||    |||||||||||| |||||||      |||||  | |||
Db       1900 TTCGTGTTCCCCGGGCAGGGCACGCAGTGGGCCGGCATGGGTGCCGAACTGCTGGACTCT 1959

Qy         64 TCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGAC 123
              ||  ||||||||||    |   |||   |  |||||||  || ||  | ||||  || |||
Db       1960 TCCGCGGTGTTCGCGGCGGCCATGGCCGAATGCGAGGCCGCACTCTCCCCGTACGTCGAC 2019

Qy        124 TGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGAT 183
              |||||  ||| || | || | || | || | || |  |   |  ||||||||  ||||
Db       2020 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 2079

Qy        184 GTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTAC 243
              || || |||||||||  ||||  |||||||||  ||  |||||     ||||||     | ||
Db       2080 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 2139

Qy        244 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 303
              ||  |   ||| | |||||| | ||||| || ||||||||||||||| |||||| | ||
Db       2140 GGGGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 2199

Qy        304 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 363
              || || |||||||||| | ||||| ||         ||  | |||||||||  | |  |
Db       2200 GCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGACCCTGCGCAGCAAGTCCATC 2259

Qy        364 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              |||||   |   |||  ||||||||||   ||  |  |||||   |   | |  ||
Db       2260 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAGCGAGGACGCCGTC 2319

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              | |||  | |  |      ||   |  || |||  || || || || || ||||| |||
Db       2320 CTGGAGCGACTGGCCG------GGTTCGACGGGCTGTCCGTCGCCGCTGTGAACGGGCCC 2373

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              | ||||||  ||||| || ||  |||     | | || ||| |      |    ||||
Db       2374 ACCGCCACCGTGGTCTCCGGTGACCCCGTACAGATCGAAGAGCTTGCTCGGGCGTGTGAG 2433

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              | |  ||||| || || |||   || || |||||||| || || ||| |||  || ||
Db       2434 GCCGATGGGGTCCGTGCGCGGGTCATTCCCGTCGACTACGCGTCCCACAGCCGGCAGGTC 2493

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              ||  | |  |  ||| |   ||| | || |  |  ||||||||||| |  |  ||
Db       2494 GAGATCATCGAGAGCGAGCTCGCCGAGGTCCTCGCCGGGCTCAGCCCGCAGGCTCCGCGC 2553

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              ||||||||||||| ||  | || ||| ||||| |   |    |    || |||| || |
Db       2554 GTGCCGTTCTTCTCGACACTCGAAGGCGCCTGGATCACCGAGCCCGTGCTCGACGGCGGC 2613

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGG-- 778
              |||||||||||||||||| ||| |  ||  | |||  |   ||||| || ||||||
Db       2614 TACTGGTACCGCAACCTGCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAGACCCTGGCC 2673

Qy        779 -CGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
               | || ||  ||  |  |  |||||||| |||||  |||||||  ||||  |   |||
Db       2674 ACCGACGAGGGCTTCACCCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTCACCATGGCC 2733

Qy        838 ATCGAAGACACC 849
              ||   |  |||
Db       2734 CTCCCCGGGACC 2745


RESULT 7
US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-5

  Query Match             34.5%;  Score 318;  DB 9;  Length 36778;
  Best Local Similarity   63.4%;  Pred. No. 4.4e-63;
  Matches  540;  Conservative    0;  Mismatches  300;  Indels   12;  Gaps    3;

Qy          4 TTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCC 63
              |||||||| |||||  |||||    |||||||||||| |||||||      |||||  | |||
Db       3641 TTCGTGTTCCCCGGGCAGGGCACGCAGTGGGCCGGCATGGGTGCCGAACTGCTGGACTCT 3700

Qy         64 TCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGAC 123
              ||  ||||||||||    |   |||   |  |||||||  || ||  | ||||  || |||
Db       3701 TCCGCGGTGTTCGCGGCGGCCATGGCCGAATGCGAGGCCGCACTCTCCCCGTACGTCGAC 3760

Qy        124 TGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGAT 183
              |||||  ||| || | || | || | || | || |  |   |  ||||||||  ||||
Db       3761 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 3820

Qy        184 GTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTAC 243
              || || |||||||||  ||||  |||||||||  ||  |||||     ||||||     | ||
Db       3821 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 3880

Qy        244 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 303
              ||  |   ||| | |||||| | ||||| || ||||||||||||||| |||||| | ||
Db       3881 GGGGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 3940

Qy        304 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 363
              || || |||||||||| | ||||| ||         ||  | |||||||||  | |  |
Db       3941 GCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGACCCTGCGCAGCAAGTCCATC 4000

Qy        364 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              |||||   |   |||  ||||||||||   ||  |  |||||   |   | |  ||
Db       4001 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAGCGAGGACGCCGTC 4060

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              | |||  | |  |      ||   |  || |||  || || || || || ||||| |||
```

```
Db       1900 TTCGTGTTCCCCGGGCAGGGCACGCAGTGGGCCGGCATGGGTGCCGAACTGCTGGACTCT 1959

Qy         64 TCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGAC 123

Db       1960 TCCGCGGTGTTCGCGGCGGCCATGGCCGAATGCGAGGCCGCACTCTCCCCGTACGTCGAC 2019

Qy        124 TGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGAT 183

Db       2020 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 2079

Qy        184 GTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTAC 243

Db       2080 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 2139

Qy        244 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 303

Db       2140 GGGGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 2199

Qy        304 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 363

Db       2200 GCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGACCCTGCGCAGCAAGTCCATC 2259

Qy        364 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420

Db       2260 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAGCGAGGACGCCGTC 2319

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480

Db       2320 CTGGAGCGACTGGCCG------GGTTCGACGGGCTGTCCGTCGCCGCTGTGAACGGGCCC 2373

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540

Db       2374 ACCGCCACCGTGGTCTCCGGTGACCCCGTACAGATCGAAGAGCTTGCTCGGGCGTGTGAG 2433

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600

Db       2434 GCCGATGGGGTCCGTGCGCGGGTCATTCCCGTCGACTACGCGTCCCACAGCCGGCAGGTC 2493

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660

Db       2494 GAGATCATCGAGAGCGAGCTCGCCGAGGTCCTCGCCGGGCTCAGCCCGCAGGCTCCGCGC 2553

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720

Db       2554 GTGCCGTTCTTCTCGACACTCGAAGGCGCCTGGATCACCGAGCCCGTGCTCGACGGCGGC 2613

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGG-- 778

Db       2614 TACTGGTACCGCAACCTGCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAGACCCTGGCC 2673

Qy        779 -CGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837

Db       2674 ACCGACGAGGGCTTCACCCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTCACCATGGCC 2733

Qy        838 ATCGAAGACACC 849

Db       2734 CTCCCCGGGACC 2745


RESULT 5
US-09-836-821-30
; Sequence 30, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-30

  Query Match             34.5%;  Score 318;  DB 9;  Length 13842;
  Best Local Similarity   63.4%;  Pred. No. 4.9e-63;
  Matches  540;  Conservative    0;  Mismatches  300;  Indels   12;  Gaps    3;

Qy          4 TTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCC 63

Db       1900 TTCGTGTTCCCCGGGCAGGGCACGCAGTGGGCCGGCATGGGTGCCGAACTGCTGGACTCT 1959

Qy         64 TCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGAC 123

Db       1960 TCCGCGGTGTTCGCGGCGGCCATGGCCGAATGCGAGGCCGCACTCTCCCCGTACGTCGAC 2019

Qy        124 TGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGAT 183

Db       2020 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 2079

Qy        184 GTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTAC 243

Db       2080 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 2139

Qy        244 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 303

Db       2140 GGGGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 2199

Qy        304 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 363

Db       2200 GCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGACCCTGCGCAGCAAGTCCATC 2259

Qy        364 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420

Db       2260 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAGCGAGGACGCCGTC 2319

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480

Db       2320 CTGGAGCGACTGGCCG------GGTTCGACGGGCTGTCCGTCGCCGCTGTGAACGGGCCC 2373

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540

Db       2374 ACCGCCACCGTGGTCTCCGGTGACCCCGTACAGATCGAAGAGCTTGCTCGGGCGTGTGAG 2433

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600

Db       2434 GCCGATGGGGTCCGTGCGCGGGTCATTCCCGTCGACTACGCGTCCCACAGCCGGCAGGTC 2493

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660

Db       2494 GAGATCATCGAGAGCGAGCTCGCCGAGGTCCTCGCCGGGCTCAGCCCGCAGGCTCCGCGC 2553

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720

Db       2554 GTGCCGTTCTTCTCGACACTCGAAGGCGCCTGGATCACCGAGCCCGTGCTCGACGGCGGC 2613

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGG-- 778

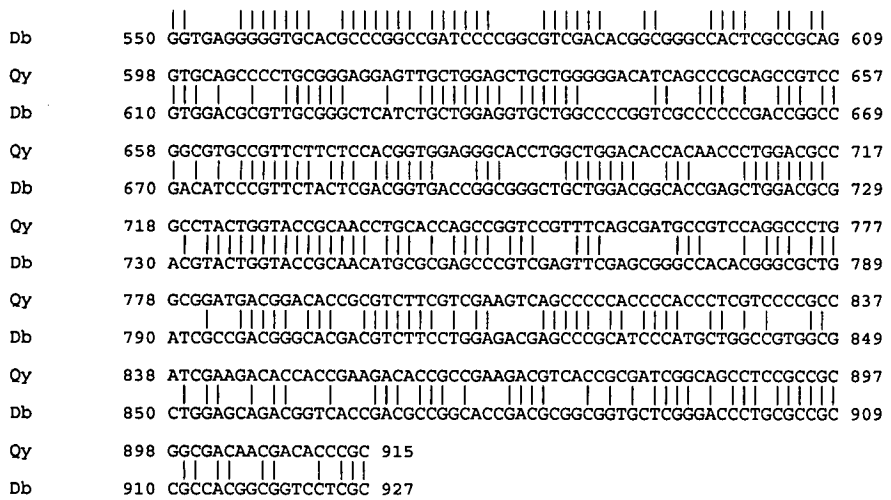
Db       2614 TACTGGTACCGCAACCTGCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAGACCCTGGCC 2673

Qy        779 -CGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837

Db       2674 ACCGACGAGGGCTTCACCCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTCACCATGGCC 2733

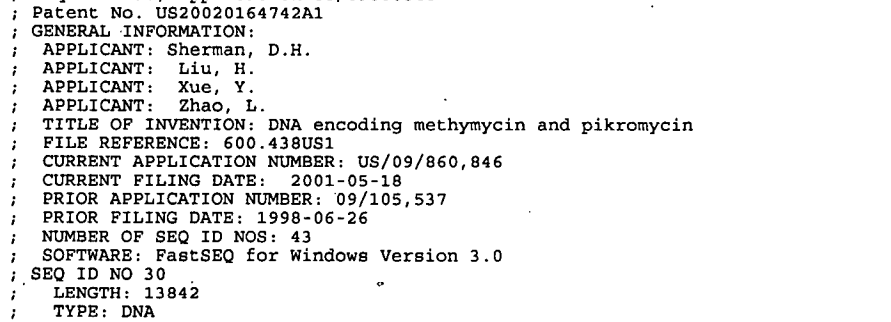
Qy        838 ATCGAAGACACC 849

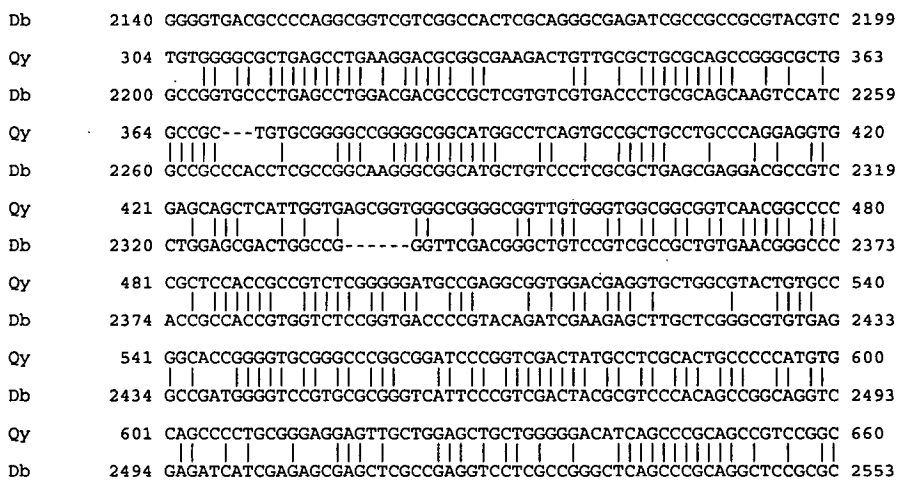
Db       2734 CTCCCCGGGACC 2745
```



RESULT 3
US-09-860-846-30
; Sequence 30, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-30

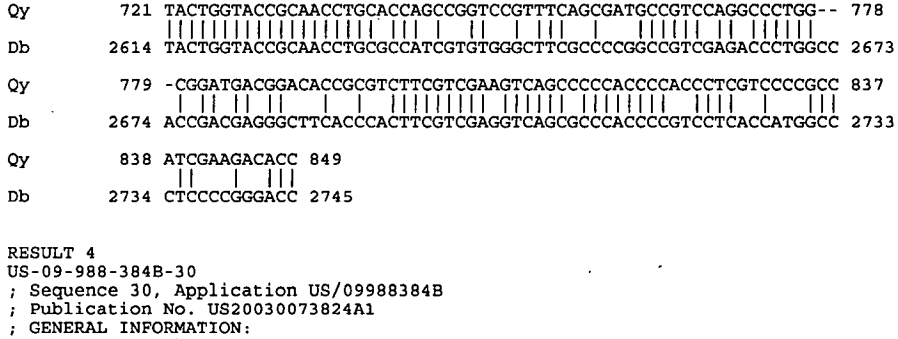
Query Match 34.5%; Score 318; DB 9; Length 13842;
Best Local Similarity 63.4%; Pred. No. 4.9e-63;
Matches 540; Conservative 0; Mismatches 300; Indels 12; Gaps 3;





RESULT 4
US-09-988-384B-30
; Sequence 30, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-30

Query Match 34.5%; Score 318; DB 9; Length 13842;
Best Local Similarity 63.4%; Pred. No. 4.9e-63;
Matches 540; Conservative 0; Mismatches 300; Indels 12; Gaps 3;



```
Db      24765 GGTTGGTCGCTGGTCGAGGTGTTGCGC---GGTGGTGGTGAGGCTGTTCTTGGGCGGGTT 24821

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240

Db      24822 GATGTGGTGCAGCCGGCGTTGTGGGCGGTGATGGTGTCACTGGCCCGGACCTGGCGGTAT 24881

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300

Db      24882 TACGGTGTGGAGCCTGCTGCGGTTGTGGGGCATTCGCAGGGTGAGATCGCTGCGGCTTGT 24941

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360

Db      24942 GTGGCTGGGGGGTTGAGTCTGGCCGATGGTGCGCGGGTGGTGGTGTTGCGGAGCCGGGCG 25001

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420

Db      25002 ATCGCCCGGATCGCTGGTGGGGGCGGCATGGTCTCGGTCGGTCTTTCAGC---TGAGCGT 25058

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480

Db      25059 GTCCGCACCATGCTCGACACCTACGGTGGCCGGGTTTCGGTCGCGGCGGTCAATGGCCCG 25118

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540

Db      25119 TCCTCGACCGTCGTGTCCGGTGACGTCCAGGCCCTGGATGAGTTGTTGGCCGGTTGTGAG 25178

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600

Db      25179 CGGGAGGGTGTCCGGGCTCGTCGTGTCCCGGTGGACTATGCCTCCCACTCCGCGCAGATG 25238

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660

Db      25239 GACCAGTTACGCGATGAGCTGCTGGAAGCGCTGGCGGACATCACTCCGCAACATTCCAGT 25298

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720

Db      25299 GTTCCGTTCTTCTCGACGGTGACGGCGGACTGGCTGGACACGACCGCTCTGGATGCGGGG 25358

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780

Db      25359 TACTGGTTCACGAATCTGCGGGAGACGGTCCGGTTCCAGGAAGCCGTCGAAGGGCTCGTG 25418

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840

Db      25419 GCTCAGGGGATGGGCGCGTTCGTCGAGTGCAGCCCGCACCCCGTCCTCGTCCCCGGTATC 25478

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900

Db      25479 GAGCAGACCCTCGACGCCCTCGACCAGAACGCCGCCGTACTCGGCTCCCTGCGGCGTGAC 25538

Qy        901 GACAACGACACCCGCCGCTTC 921

Db      25539 GAAGGCGGCCTGGACCGACTC 25559


RESULT 2
US-09-735-056-29
; Sequence 29, Application US/09735056
; Publication No. US20030013662A1
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/735,056
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/858,003
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1010 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-735-056-29

  Query Match             39.4%;  Score 362.8;  DB 9;  Length 1010;
  Best Local Similarity   63.7%;  Pred. No. 6.1e-73;
  Matches  585;  Conservative    0;  Mismatches  327;  Indels    6;  Gaps    2;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60

Db         13 GTGTTCGTGTTCCCCGGCCAGGGCTCGCAGTGGGCCGGAATGGCCGAGGGGCTGCTGGAG 72

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120

Db         73 CGGTCCGGCGCGTTCCGGAGTGCGGCCGACTCGTGCGACGCCGCGCTGCGGCCGTACCTC 132

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180

Db        133 GGCTGGTCGGTGCTGAGCGTGCTGCGCGGGGAACCGGACGCGCCCTCGCTCGACCGGGTC 192

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240

Db        193 GACGTCGTGCAGCCGGTGCTGTTCACGATGATGGTCTCGCTCGCGGCGGTCTGGCGTGCG 252

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300

Db        253 CTGGGGGTGGAACCGGCGGCGGTCGTCGGGCACTCGCAGGGTGAGATCGCCGCTGCCCAT 312

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360

Db        313 GTCGCCGGTGCGCTGTCGCTGGACGACTCGGCCCGGATCGTCGCCCTGCGCAGTCGGGCG 372

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420

Db        373 TGGCTCGGACTGGCGGGCAAGGGCGGCATGGTGGCGGTGCCGATGCCGGCGGAGGAGCTG 432

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480

Db        433 CGGCCGCGGCT---GGTGACGTGGGGGGACCGTCTGGCCGTCGCCGCCGTCAACAGCCCC 489

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540

Db        490 GGTTCCTGCGCCGTCGCAGGCGACCCGGAGGCGCTGGCCGAACTGGTGGCGCTGCTGACC 549

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCC---GGTCGACTATGCCTCGCACTGCCCCCAT 597
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 17:56:34 ; Search time 108.842 Seconds
(without alignments)
12253.192 Million cell updates/sec

Title: US-09-914-286-1_COPY_3148_4068
Perfect score: 921
Sequence: 1 gtcttcgtgtttcccggaca.....acaacgacacccgccgcttc 921

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	373.8	40.6	50937	9	US-09-808-880-1	Sequence 1, Appli
2	362.8	39.4	1010	9	US-09-735-056-29	Sequence 29, Appl
3	318	34.5	13842	9	US-09-860-846-30	Sequence 30, Appl
4	318	34.5	13842	9	US-09-988-384B-30	Sequence 30, Appl
5	318	34.5	13842	9	US-09-836-821-30	Sequence 30, Appl
6	318	34.5	13842	10	US-09-861-289-30	Sequence 30, Appl
7	318	34.5	36778	9	US-09-860-846-5	Sequence 5, Appli
8	318	34.5	36778	9	US-09-836-821-5	Sequence 5, Appli
9	318	34.5	36778	10	US-09-861-289-5	Sequence 5, Appli
10	318	34.5	37948	9	US-09-988-384B-5	Sequence 5, Appli
11	318	34.5	38506	9	US-09-793-708-19	Sequence 19, Appl
12	311	33.8	11220	9	US-09-860-846-32	Sequence 32, Appl
13	311	33.8	11220	9	US-09-988-384B-32	Sequence 32, Appl
14	311	33.8	11220	9	US-09-836-821-32	Sequence 32, Appl
15	311	33.8	11220	10	US-09-861-289-32	Sequence 32, Appl
16	308.6	33.5	1035	9	US-09-735-056-30	Sequence 30, Appl
17	305.8	33.2	4689	9	US-09-860-846-34	Sequence 34, Appl
18	305.8	33.2	4689	9	US-09-988-384B-34	Sequence 34, Appl
19	305.8	33.2	4689	9	US-09-836-821-34	Sequence 34, Appl
20	305.8	33.2	4689	10	US-09-861-289-34	Sequence 34, Appl
21	298.4	32.4	68750	9	US-10-014-717-1	Sequence 1, Appli
22	280.6	30.5	4041	9	US-09-860-846-36	Sequence 36, Appl
23	280.6	30.5	4041	9	US-09-988-384B-36	Sequence 36, Appl
24	280.6	30.5	4041	9	US-09-836-821-36	Sequence 36, Appl
25	280.6	30.5	4041	10	US-09-861-289-36	Sequence 36, Appl
26	187.4	20.3	15872	9	US-09-860-846-1	Sequence 1, Appli
27	187.4	20.3	15872	9	US-09-988-384B-1	Sequence 1, Appli
28	187.4	20.3	15872	9	US-09-836-821-1	Sequence 1, Appli
29	187.4	20.3	15872	10	US-09-861-289-1	Sequence 1, Appli
30	164.4	17.9	5484	9	US-09-712-363-115	Sequence 115, App
31	152.8	16.6	1030	9	US-09-735-056-2	Sequence 2, Appli
32	147.4	16.0	4851	9	US-09-712-363-116	Sequence 116, App
33	147	16.0	4209	9	US-09-712-363-20	Sequence 20, Appl
34	144.6	15.7	426	9	US-10-125-815-5	Sequence 5, Appli
35	138.2	15.0	426	9	US-10-125-815-6	Sequence 6, Appli
36	64.4	7.0	3331	9	US-09-373-658-31	Sequence 31, Appl
37	62.2	6.8	925	9	US-09-735-056-1	Sequence 1, Appli
38	61.8	6.7	12733	9	US-10-032-393-47	Sequence 47, Appl
39	61.8	6.7	12739	9	US-10-032-393-8	Sequence 8, Appli
40	60.4	6.6	88421	9	US-09-976-059-1	Sequence 1, Appli
c 41	58.6	6.4	3372	9	US-10-067-457-2	Sequence 2, Appli
c 42	58.6	6.4	38186	9	US-09-373-658-38	Sequence 38, Appl
c 43	57.8	6.3	2307	9	US-09-893-519A-87	Sequence 87, Appl
44	57.2	6.2	4257	9	US-09-825-288A-1	Sequence 1, Appli
c 45	57	6.2	520	9	US-10-184-644-332	Sequence 332, App

ALIGNMENTS

RESULT 1
US-09-808-880-1
; Sequence 1, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/808,880
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/428,517
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/106,100
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artifical Sequence: Recombinant DNA
US-09-808-880-1

Query Match 40.6%; Score 373.8; DB 9; Length 50937;
Best Local Similarity 64.4%; Pred. No. 1.2e-75;
Matches 593; Conservative 0; Mismatches 322; Indels 6; Gaps 2;

```
Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              || ||||| ||||| || ||||||  |||||||  |||||||  |    |  |||| |
Db      24645 GTGTTCGTCTTTCCTGGTCAGGGTTCGCAGTGGGTGGGGATGGCGGCTGGGTTGCTGGAT 24704

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
                | | ||||||||||||   |  | ||  || ||| |  | ||  ||||  |||  |  |
Db      24705 GCGTGTCCGGTGTTCGCGGAGGCGGTGGCGGAGTGTGCCGCGGTGCTGGATCCGGTGACG 24764

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
```

```
Db      21250 GCACGGCGCCT---CCAGCGCTATGGCGATCGGCTCTCCATCGGGGCGATCAACAGCCCT 21306

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540

Db      21307 CGTTTCACGACGATCTCCGGCGAGCCCCCTGCCGTCGCCGCCCTGCTCCGCGATCTGGAG 21366

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600

Db      21367 TCCGAGGGCGTCTTCGCCCTCAAGCTGAGTTACGACTTCGCCTCCCACTCCGCGCAGGTC 21426

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660

Db      21427 GAGTCGATTCGCGACGAGCTCCTCGATCTCCTGTCGTGGCTCGAGCCGCGCTCGACGGCG 21486

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720

Db      21487 GTCCCGTTCTACTCCACGGTGAGCGGCGCCGCGATCGACGGGAGCGAGCTCGACGCCGCC 21546

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780

Db      21547 TACTGGTACCGGAACCTCCGGCAGCCGGTCCGCTTCGCAGACGCTGTGCAAGGCCTCCTT 21606

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840

Db      21607 GCCGGAGAACATCGCTTCTTCGTGGAGGTGAGCCCCAGTCCTGTGCTGACCTTGGCCTTG 21666

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900

Db      21667 CACGAGCTCCTCGAAGCGTCGGAGCGCTCGGCGGCGGTGGTCGGCTCTCTGTGGAGCGAC 21726

Qy        901 GACAACGACACCCGCCGCTTC 921

Db      21727 GAAGGGGATCTACGGCGCTTC 21747
```

Search completed: June 18, 2003, 01:06:54
Job time : 51.8967 secs

```
                ||||||||  || | |   | || ||| | || | ||| |   |   |   || |||| |
Db      20950 GACTGGTCGCTGCTCGCGGTGCTCCGCGGCGAGGAGGGCGCGCCCCCGCTCGACCGGGTC 21009

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
                || |||||||||||| | |||||||     | |||| ||  |||| || |||||||| |||
Db      21010 GACGTGGTCCAGCCCGCGCTGTTCTCGATGATGGTCTCGCTGGCCGCCCTGTGGCGCTCC 21069

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
                ||  |||| ||||||||||||  | ||||||  |||||||||||||||| || ||   |
Db      21070 ATGGGCGTCGAGCCCGACGCGGTGGTCGGCCATAGCCAGGGCGAGATCGCCGCGGCCTGT 21129

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
                |||   || ||||||   ||  ||||||| || |||   || ||||||||||||| |||
Db      21130 GTGGCGGGCGCGCTGTCGCTCGAGGACGCTGCCAAGCTGGTGGCGCTGCGCAGCCGTGCG 21189

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
                || |  |   |    |||| ||| | ||||||| | |||  |||||| |     |||||
Db      21190 CTCGTGGAGCTCGCCGGCCAGGGGGCCATGGCCGCGGTGGAGCTGCCGGAGGCCGAGGTC 21249

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
                |  | || | |    |||| |  |  |  ||| | |   | | |||| ||||| ||||
Db      21250 GCACGGCGCCT---CCAGCGCTATGGCGATCGGCTCTCCATCGGGGCGATCAACAGCCCT 21306

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
                || | |||  |  |||| || ||  ||  || | ||  ||||    |   |
Db      21307 CGTTTCACGACGATCTCCGGCGAGCCCCCTGCCGTCGCCGCCCTGCTCCGCGATCTGGAG 21366

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
                |   || ||   ||||   | |     |||||  ||||| |||| | | || ||
Db      21367 TCCGAGGGCGTCTTCGCCCTCAAGCTGAGTTACGACTTCGCCTCCCACTCCGCGCAGGTC 21426

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
                || |  | || || ||| | || || || ||| |   ||   ||||   || | |
Db      21427 GAGTCGATTCGCGACGAGCTCCTCGATCTCCTGTCGTGGCTCGAGCCGCGCTCGACGGCG 21486

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
                || ||||||| ||||||||||   ||| ||  | | |||   |     || |||||||||
Db      21487 GTCCCGTTCTACTCCACGGTGAGCGGCGCCGCGATCGACGGGAGCGAGCTCGACGCCGCC 21546

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
                |||||||||||  ||||| |  |||||||||||  |||   || || ||  || | |||
Db      21547 TACTGGTACCGGAACCTCCGGCAGCCGGTCCGCTTCGCAGACGCTGTGCAAGGCCTCCTT 21606

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
                | .|  | ||| ||| |||||||  || || ||||||   ||    ||    |  ||| |
Db      21607 GCCGGAGAACATCGCTTCTTCGTGGAGGTGAGCCCCAGTCCTGTGCTGACCTTGGCCTTG 21666

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
                | ||   |  ||||   | |          |   | | | ||||   ||  |  ||| |
Db      21667 CACGAGCTCCTCGAAGCGTCGGAGCGCTCGGCGGCGGTGGTCGGCTCTCTGTGGAGCGAC 21726

Qy        901 GACAACGACACCCGCCGCTTC 921
                ||    ||    || ||||||
Db      21727 GAAGGGGATCTACGGCGCTTC 21747


RESULT 15
US-08-458-076A-6
; Sequence 6, Application US/08458076A
; Patent No. 5698425
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,076A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-458-076A-6

  Query Match           37.5%;  Score 345;  DB 1;  Length 28958;
  Best Local Similarity   61.7%;  Pred. No. 9.9e-52;
  Matches  568;  Conservative    0;  Mismatches  350;  Indels    3;  Gaps    1;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
                |||||||| |||||  || || ||   ||||||||  |||||||    |    |||||||
Db      20830 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAT 20889

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
                ||||  |||||  |||       | | |  | |||||||||    ||  ||  |||||     |||
Db      20890 ACCTCGCCGGTCTTCCGGGCACAGCTCGAAGCGTGCGAGCGCGCCCTCGCGCCCCCACGTG 20949

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
                ||||||||  || | |   | || ||| | || | ||| |   |   |   || |||| |
Db      20950 GACTGGTCGCTGCTCGCGGTGCTCCGCGGCGAGGAGGGCGCGCCCCCGCTCGACCGGGTC 21009

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
                || |||||||||||| | |||||||     | ||||| ||  |||| || |||||||| |||
Db      21010 GACGTGGTCCAGCCCGCGCTGTTCTCGATGATGGTCTCGCTGGCCGCCCTGTGGCGCTCC 21069

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
                ||  |||| ||||||||||||  | ||||||  |||||||||||||||| || ||   |
Db      21070 ATGGGCGTCGAGCCCGACGCGGTGGTCGGCCATAGCCAGGGCGAGATCGCCGCGGCCTGT 21129

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
                |||   || ||||||   ||  ||||||| || |||   || ||||||||||||| |||
Db      21130 GTGGCGGGCGCGCTGTCGCTCGAGGACGCTGCCAAGCTGGTGGCGCTGCGCAGCCGTGCG 21189

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
                || |  |   |    |||| ||| | ||||||| | |||  |||||| |     |||||
Db      21190 CTCGTGGAGCTCGCCGGCCAGGGGGCCATGGCCGCGGTGGAGCTGCCGGAGGCCGAGGTC 21249

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
                |  | || | |    |||| |  |  |  ||| | |   | | |||| ||||| ||||
```

```
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-342-6

  Query Match             37.5%;  Score 345;  DB 1;  Length 28958;
  Best Local Similarity   61.7%;  Pred. No. 9.9e-52;
  Matches  568;  Conservative    0;  Mismatches  350;  Indels    3;  Gaps    1;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              ||||||||  |||||  ||  ||  ||    ||||||||  |||||||     |     |||||||
Db      20830 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAT 20889

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
               ||||  |||||  |||       | |  |  | |||||||||    ||  ||  |||||     |||
Db      20890 ACCTCGCCGGTCTTCCGGGCACAGCTCGAAGCGTGCGAGCGCGCCCTCGCGCCCCACGTG 20949

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||  ||  | |    | || ||| | || | ||| |    |    |   || |||| |
Db      20950 GACTGGTCGCTGCTCGCGGTGCTCCGCGGCGAGGAGGGCGCGCCCCCGCTCGACCGGGTC 21009

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              || ||||||||||| | ||||||||     | |||||  ||   |||| ||  |||||||| |||
Db      21010 GACGTGGTCCAGCCCGCGCTGTTCTCGATGATGGTCTCGCTGGCCGCCCTGTGGCGCTCC 21069

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
                ||  |||| ||||||||||   | ||||||  |||||||||||||| || ||    |
Db      21070 ATGGGCGTCGAGCCCGACGCGGTGGTCGGCCATAGCCAGGGCGAGATCGCCGCGGCCTGT 21129

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              |||    || ||||||   ||  ||||||| || |||   || ||||||||||||| |||
Db      21130 GTGGCGGGCGCGCTGTCGCTCGAGGACGCTGCCAAGCTGGTGGCGCTGCGCAGCCGTGCG 21189

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              || |   |    |    |||| ||| | ||||||| | |||  ||||||  |      |||||
Db      21190 CTCGTGGAGCTCGCCGGCCAGGGGGCCATGGCCGCGGTGGAGCTGCCGGAGGCCGAGGTC 21249

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              |  | ||  |  |     |||| |  |  |  ||| | |   | | |||| ||||| ||||
Db      21250 GCACGGCGCCT---CCAGCGCTATGGCGATCGGCTCTCCATCGGGGCGATCAACAGCCCT 21306

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              || | |||  |  |||| || ||  ||   || || | ||   ||||     |    |
Db      21307 CGTTTCACGACGATCTCCGGCGAGCCCCCTGCCGTCGCCGCCCTGCTCCGCGATCTGGAG 21366

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              |   || ||   ||||     | |       |||||  |||||  |||| | | || ||
Db      21367 TCCGAGGGCGTCTTCGCCCTCAAGCTGAGTTACGACTTCGCCTCCCACTCCGCGCAGGTC 21426

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              || |  | || || ||| | || || || |||  |     ||   ||||   || | |
Db      21427 GAGTCGATTCGCGACGAGCTCCTCGATCTCCTGTCGTGGCTCGAGCCGCGCTCGACGGCG 21486

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              || ||||||| ||||||||||   ||| ||  | | |||    |    || |||||||||
Db      21487 GTCCCGTTCTACTCCACGGTGAGCGGCGCCGCGATCGACGGGAGCGAGCTCGACGCCGCC 21546

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              |||||||||| |||||  |  |||||||||| |||   || || || || | |||
Db      21547 TACTGGTACCGGAACCTCCGGCAGCCGGTCCGCTTCGCAGACGCTGTGCAAGGCCTCCTT 21606

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              |  |  | ||| ||| |||||||| || || ||||||   ||     ||    | ||| |
Db      21607 GCCGGAGAACATCGCTTCTTCGTGGAGGTGAGCCCCAGTCCTGTGCTGACCTTGGCCTTG 21666

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
              | ||   |  |||||   | |       |   | | | |||||   ||  |  ||| |
Db      21667 CACGAGCTCCTCGAAGCGTCGGAGCGCTCGGCGGCGGTGGTCGGCTCTCTGTGGAGCGAC 21726

Qy        901 GACAACGACACCCGCCGCTTC 921


              ||    ||    || ||||||
Db      21727 GAAGGGGATCTACGGCGCTTC 21747


RESULT 14
US-08-457-646A-6
; Sequence 6, Application US/08457646A
; Patent No. 5679560
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,646A
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-457-646A-6

  Query Match             37.5%;  Score 345;  DB 1;  Length 28958;
  Best Local Similarity   61.7%;  Pred. No. 9.9e-52;
  Matches  568;  Conservative    0;  Mismatches  350;  Indels    3;  Gaps    1;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              ||||||||  |||||  ||  ||  ||    ||||||||  |||||||     |     |||||||
Db      20830 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAT 20889

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
               ||||  |||||  |||       | |  |  | |||||||||    ||  ||  |||||     |||
Db      20890 ACCTCGCCGGTCTTCCGGGCACAGCTCGAAGCGTGCGAGCGCGCCCTCGCGCCCCACGTG 20949

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
```

```
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/456,837
;       FILING DATE:  01-JUN-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/457,205
;       FILING DATE:  01-JUN-1995
;       APPLICATION NUMBER:  08/258,261
;       FILING DATE:  08-Jun-1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Elmer, James Scott
;       REGISTRATION NUMBER:  36,129
;       REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  919-541-8614
;       TELEFAX:  919-541-8689
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  28958 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
US-08-456-837-6

  Query Match             37.5%;  Score 345;  DB 1;  Length 28958;
  Best Local Similarity   61.7%;  Pred. No. 9.9e-52;
  Matches  568;  Conservative    0;  Mismatches  350;  Indels    3;  Gaps    1;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              ||||||| ||||| || || ||   |||||||  |||||||   |    |||||||
Db      20830 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAT 20889

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
               |||| ||||| |||      | | |  | |||||||||   || || |||||    |||
Db      20890 ACCTCGCCGGTCTTCCGGGCACAGCTCGAAGCGTGCGAGCGCGCCCTCGCGCCCCACGTG 20949

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||  || | |    | || ||| | || | ||| |    |    |   || |||| |
Db      20950 GACTGGTCGCTGCTCGCGGTGCTCCGCGGCGAGGAGGGCGCGCCCCCGCTCGACCGGGTC 21009

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              || |||||||||| | |||||||    | ||||| ||  |||| || |||||||| |||
Db      21010 GACGTGGTCCAGCCCGCGCTGTTCTCGATGATGGTCTCGCTGGCCGCCCTGTGGCGCTCC 21069

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
                ||  |||| ||||||||||  | ||||||  ||||||||||||||| || ||   |
Db      21070 ATGGGCGTCGAGCCCGACGCGGTGGTCGGCCATAGCCAGGGCGAGATCGCCGCGGCCTGT 21129

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              |||   || ||||||   ||  ||||||| || |||   || |||||||||||||| |||
Db      21130 GTGGCGGGCGCGCTGTCGCTCGAGGACGCTGCCAAGCTGGTGGCGCTGCGCAGCCGTGCG 21189

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              || |  |   |   |||| ||| | ||||||| | |||  ||||||  |     |||||
Db      21190 CTCGTGGAGCTCGCCGGCCAGGGGGCCATGGCCGCGGTGGAGCTGCCGGAGGCCGAGGTC 21249

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              |  | || | |     |||| |  |  |  ||| | |   | | |||| ||||| ||||
Db      21250 GCACGGCGCCT---CCAGCGCTATGGCGATCGGCTCTCCATCGGGGCGATCAACAGCCCT 21306

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              || | |||  |  |||| || ||  ||  || || | ||  ||||    |  |  |
Db      21307 CGTTTCACGACGATCTCCGGCGAGCCCCCTGCCGTCGCCGCCCTGCTCCGCGATCTGGAG 21366

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
                |   || ||    ||||    | |       |||||  ||||| |||| | | || ||
Db      21367 TCCGAGGGCGTCTTCGCCCTCAAGCTGAGTTACGACTTCGCCTCCCACTCCGCGCAGGTC 21426

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              || |  | || || ||| | || || || |||  |     ||    ||||   || | |
Db      21427 GAGTCGATTCGCGACGAGCTCCTCGATCTCCTGTCGTGGCTCGAGCCGCGCTCGACGGCG 21486

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              || ||||||| ||||||||||   ||| ||  | | |||   |     || |||||||||
Db      21487 GTCCCGTTCTACTCCACGGTGAGCGGCGCCGCGATCGACGGGAGCGAGCTCGACGCCGCC 21546

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              |||||||||| ||||| |  |||||||||| |||   || || || || | |||
Db      21547 TACTGGTACCGGAACCTCCGGCAGCCGGTCCGCTTCGCAGACGCTGTGCAAGGCCTCCTT 21606

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              |  |  | ||| ||| ||||||| || || ||||||   ||    ||   |  ||| |
Db      21607 GCCGGAGAACATCGCTTCTTCGTGGAGGTGAGCCCCAGTCCTGTGCTGACCTTGGCCTTG 21666

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
              | ||   | |||||   | |         |   | | | |||||   ||  |  ||| |
Db      21667 CACGAGCTCCTCGAAGCGTCGGAGCGCTCGGCGGCGGTGGTCGGCTCTCTGTGGAGCGAC 21726

Qy        901 GACAACGACACCCGCCGCTTC 921
              ||    ||    || ||||||
Db      21727 GAAGGGGATCTACGGCGCTTC 21747


RESULT 13
US-08-457-342-6
; Sequence 6, Application US/08457342
; Patent No. 5662898
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,342
;      FILING DATE:  01-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;      APPLICATION NUMBER:  08/258,261
;      FILING DATE:  08-Jun-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/258,261B
;      FILING DATE:  08-JUN-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/457,205
;      FILING DATE:  01-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Elmer, James Scott
;      REGISTRATION NUMBER:  36,129
;      REFERENCE/DOCKET NUMBER:  CGC 1506/CIP3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8614
;      TELEFAX:  919-541-8689
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28958 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-258-261B-6

  Query Match             37.5%;  Score 345;  DB 1;  Length 28958;
  Best Local Similarity   61.7%;  Pred. No. 9.9e-52;
  Matches  568;  Conservative    0;  Mismatches  350;  Indels    3;  Gaps    1;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              |||||||| |||| || || ||   |||||||  |||||||    |    |||||||
Db      20830 GTCTTCGTCTTTCCTGGGCAAGGCTCGCAGTGGGAAGGGATGGCCCTCTCCCTGCTCGAT 20889

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
                |||| |||| |||     | | |  | |||||||||   || || |||||     |||
Db      20890 ACCTCGCCGGTCTTCCGGGCACAGCTCGAAGCGTGCGAGCGCGCCCTCGCGCCCCACGTG 20949

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||  || | |   | || ||| | || | ||| | .  |    |  || |||| |
Db      20950 GACTGGTCGCTGCTCGCGGTGCTCCGCGGCGAGGAGGGCGCGCCCCCGCTCGACCGGGTC 21009

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              || |||||||||| | |||||||     | ||||| ||  |||| || |||||||| |||
Db      21010 GACGTGGTCCAGCCCGCGCTGTTCTCGATGATGGTCTCGCTGGCCGCCCTGTGGCGCTCC 21069

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
                ||  |||| ||||||||||  | ||||||  ||||||||||||||| || ||   |
Db      21070 ATGGGCGTCGAGCCCGACGCGGTGGTCGGCCATAGCCAGGGCGAGATCGCCGCGGCCTGT 21129

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              |||   || ||||||   ||  ||||||| || |||   || ||||||||||||| |||
Db      21130 GTGGCGGGCGCGCTGTCGCTCGAGGACGCTGCCAAGCTGGTGGCGCTGCGCAGCCGTGCG 21189

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              || |  |   |  . |||| ||| | ||||||| | |||  ||||||  |     |||||
Db      21190 CTCGTGGAGCTCGCCGGCCAGGGGGCCATGGCCGCGGTGGAGCTGCCGGAGGCCGAGGTC 21249

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              |  | || | |      |||| |  |  |  ||| | |   | | |||| ||||| ||||
Db      21250 GCACGGCGCCT---CCAGCGCTATGGCGATCGGCTCTCCATCGGGGCGATCAACAGCCCT 21306

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              || | |||  |  |||| || ||  ||   || || | ||   |||     |   |
Db      21307 CGTTTCACGACGATCTCCGGCGAGCCCCCTGCCGTCGCCGCCCTGCTCCGCGATCTGGAG 21366

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
                |   || ||    ||||    | |      |||||  ||||| |||| | | || ||
Db      21367 TCCGAGGGCGTCTTCGCCCTCAAGCTGAGTTACGACTTCGCCTCCCACTCCGCGCAGGTC 21426

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              || |  | || || ||| | || || || ||| |    ||  ||||   || | |
Db      21427 GAGTCGATTCGCGACGAGCTCCTCGATCTCCTGTCGTGGCTCGAGCCGCGCTCGACGGCG 21486

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              || ||||||| |||||||||   ||| ||  | | |||   |    || ||||||||
Db      21487 GTCCCGTTCTACTCCACGGTGAGCGGCGCCGCGATCGACGGGAGCGAGCTCGACGCCGCC 21546

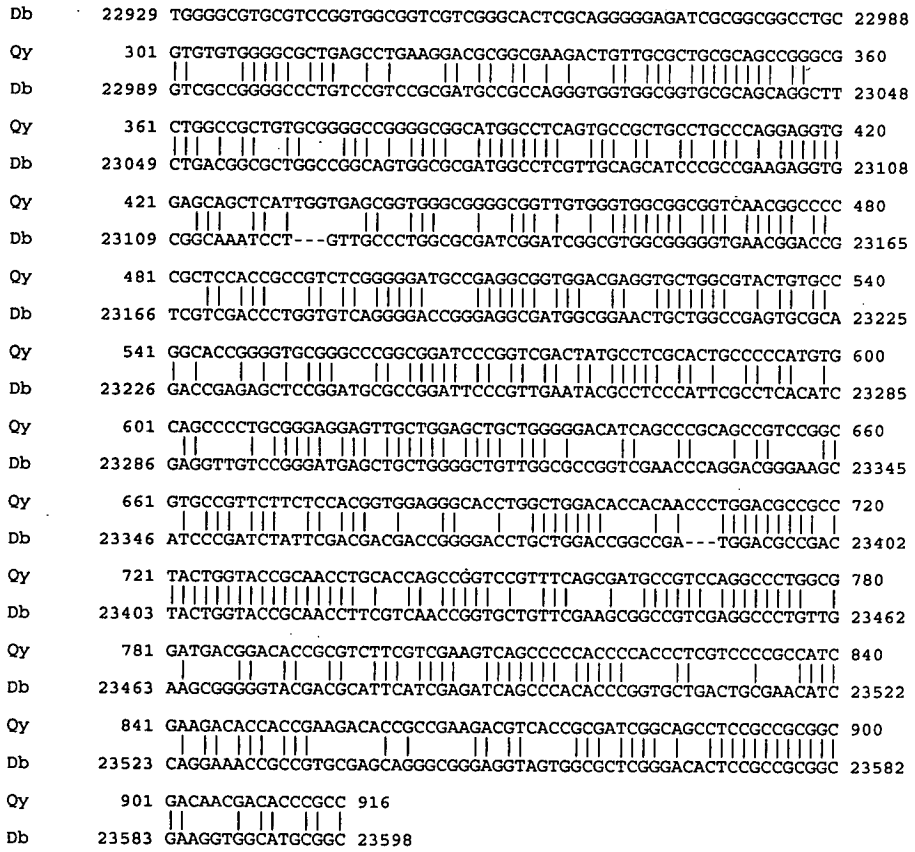
Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              |||||||||||| ||||| |  |||||||||| |||   || || || || | |||
Db      21547 TACTGGTACCGGAACCTCCGGCAGCCGGTCCGCTTCGCAGACGCTGTGCAAGGCCTCCTT 21606

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              |  |  | ||| ||| |||||| || || ||||||   ||   ||   |  ||| |
Db      21607 GCCGGAGAACATCGCTTCTTCGTGGAGGTGAGCCCCAGTCCTGTGCTGACCTTGGCCTTG 21666

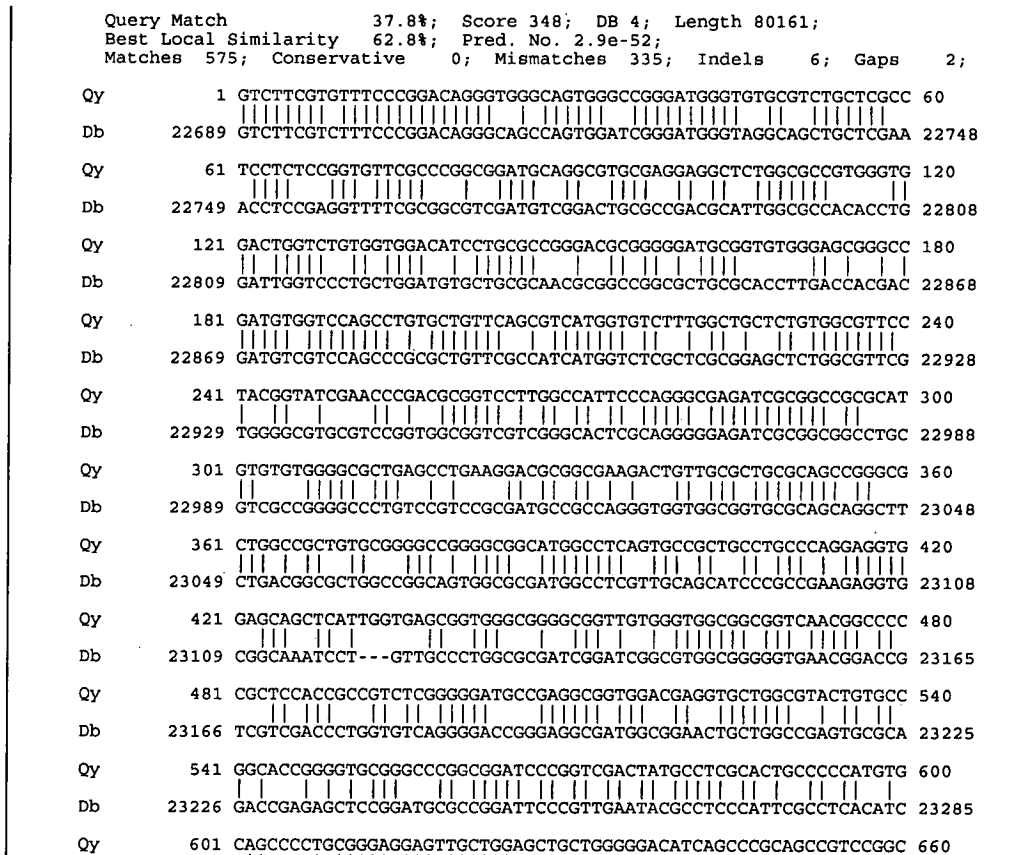
Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
              | |||   | |||||   | |         |   | | | |||||   ||  |  ||| |
Db      21667 CACGAGCTCCTCGAAGCGTCGGAGCGCTCGGCGGCGGTGGTCGGCTCTCTGTGGAGCGAC 21726

Qy        901 GACAACGACACCCGCCGCTTC 921
              ||    ||    || ||||||
Db      21727 GAAGGGGATCTACGGCGCTTC 21747


RESULT 12
US-08-456-837-6
; Sequence 6, Application US/08456837
; Patent No. 5643774
;  GENERAL INFORMATION:
;    APPLICANT:  Schupp, Thomas
;    APPLICANT:  Ligon, James M.
;    APPLICANT:  Beck, James Joseph
;    APPLICANT:  Hill, Dwight Steven
;    APPLICANT:  Ryals, John Andrew
;    APPLICANT:  Gaffney, Thomas Deane
;    APPLICANT:  Lam, Stephen Ting
;    APPLICANT:  Hammer, Phillip E.
;    APPLICANT:  Uknes, Scott Joseph
;    TITLE OF INVENTION:  Genes for the synthesis of
;    TITLE OF INVENTION:  antipathogenic substances
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ciba-Geigy Corporation
;      STREET:  7 Skyline Drive
;      CITY:  Hawthorne
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10532
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```



RESULT 10
US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 80161
; TYPE: DNA
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1



RESULT 11
US-08-258-261B-6
; Sequence 6, Application US/08258261B
; Patent No. 5639949

```
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10707..10964
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 2"
US-08-439-009A-1

  Query Match            38.0%;  Score 349.6;  DB 3;  Length 11219;
  Best Local Similarity  62.7%;  Pred. No. 1.6e-52;
  Matches  579;  Conservative    0;  Mismatches  339;  Indels    6;  Gaps    2;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              || |||||  ||  ||  ||  |||||||  |||||||  ||  |||    | |||    | |
Db       8265 GTGTTCGTCTTCCCTGGGCAGGGTGCCCAGTGGGAAGGCATG---GCGCGGGAGTTGTTG 8321

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
                |   |||  ||  ||||||   |   |||     | ||||||  | ||    ||  ||   |  ||
Db       8322 CCGGTTCCCGTCTTCGCCGAGTCGATCGCCGAGTGCGATGCGGTGTTGTCGGAGGTGGCC 8381

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              |   |  ||  |||    ||  |  |||    |        |||  |    ||  |  ||||||||  |
Db       8382 GGATTCTCGGTGTCCGAGGTGCTGGAGCCACGTCCGGACGCGCCGTCGCTGGAGCGGGTC 8441

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              || |||||  |||||  |||||||||    ||  ||||||||  ||||     ||||||||  ||
Db       8442 GACGTGGTGCAGCCGGTGCTGTTCGCGGTGATGGTGTCGCTGGCGCGGTTGTGGCGTGCC 8501

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              |  ||||   ||   ||     ||  |||  ||  ||  ||  |||||  |||||||| ||||||
Db       8502 TGCGGTGCCGTTCCTTCGGCCGTCATAGGGCACTCGCAGGGTGAGATCGCCGCCGCGGTG 8561

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              |||    ||  |||  ||    |||  ||||||      |        ||  ||  |   ||     |||||
Db       8562 GTGGCGGGAGCGTTGTCGCTGGAGGACGGCATGCGCGTCGTCGCCCGCCGGTCGAGGGCG 8621

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGC---CGCTGCCTGCCCAGGAG 417
              ||   ||  ||   |||||||||  |||||  |||  ||||       |    ||  |  |
Db       8622 GTGCGTGCGGTCGCGGGCCGGGGGAGCATGCTCTCGGTGCGCGGCGGCCGCTCCGACGTC 8681

Qy        418 GTGGAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGC 477
              | | |||  ||||    |  ||  | |||  | ||  ||  ||  |||  ||  ||||||||||||
Db       8682 GAGAAGCTGCTCGCCGACGACAGCTGGACCGGCAGGCTGGAGGTCGCCGCGGTCAACGGC 8741

Qy        478 CCCCGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGT 537
              |||  | ||    |  ||  | ||  ||  |||  ||||||  |   ||||  |  ||||  ||||||
Db       8742 CCCGACGCCGTGGTGGTGGCCGGTGACGCCCAGGCGGCGCGCGAGTTCCTGGAGTACTGC 8801

Qy        538 GCCGGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCAT 597
              |  |||    ||  |  ||  |||||   |||||||| |||||  |||||||||  |  |  ||
Db       8802 GAGGGCGTGGGCATCCGCGCCCGCGCGATCCCGGTGGACTACGCCTCGCACACCGCGCAC 8861

Qy        598 GTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCC 657
              ||  |||||  ||||  ||  ||  ||  |  ||  |||||  |  ||||  |||||  |  | ||
Db       8862 GTCGAGCCCGTGCGCGACGAACTGGTCCAGGCGCTGGCCGGGATCACCCCGCGACGGGCC 8921

Qy        658 GGCGTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCC 717
              |  ||||||||||||||||||  ||   |||  ||  ||  |||  |||     ||||||||
Db       8922 GAGGTGCCGTTCTTCTCCACCCTGACCGGCGACTTCCTCGACGGCACCGAGCTGGACGCG 8981

Qy        718 GCCTACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTG 777
              |  ||||||||||||||||||||  ||  |||||     |||  |   |||||  |||||  |||
Db       8982 GGCTACTGGTACCGCAACCTGCGTCACCCGGTGGAGTTCCACTCCGCCGTCGCAGGCGCTG 9041

Qy        778 GCGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
              |  |||  |  |||  ||        |||  ||||  |||||||||  |||||     |||  |    |
Db       9042 ACCGACCAGGGATACGCGACGTTCATCGAGGTCAGCCCGCACCCGGTGCTGGCGTCGAGC 9101

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897
              ||  |  ||  |||  |||  |||  |||      ||||    |  |  |||| |  ||     |||
Db       9102 GTCCAGGAGACCCTCGACGACGCCGAGTCGGACGCGGCGGTGCTCGGGACGCTGGAACGC 9161

Qy        898 GGCGACAACGACACCCGCCGCTTC 921
              |  ||     ||||  ||  |||||||
Db       9162 GACGCGGGCGACGCCGACCGCTTC 9185


RESULT 9
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  80161 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-036-987A-1

  Query Match            37.8%;  Score 348;  DB 3;  Length 80161;
  Best Local Similarity  62.8%;  Pred. No. 2.9e-52;
  Matches  575;  Conservative    0;  Mismatches  335;  Indels    6;  Gaps    2;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              ||||||||  ||||||||||||||  | ||||||  ||||||||||  ||  |||||||
Db      22689 GTCTTCGTCTTTCCCGGACAGGGCAGCCAGTGGATCGGGATGGGTAGGCAGCTGCTCGAA 22748

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
                ||||   ||| |||||   |  ||||  ||  ||||  ||  ||  |||||||      ||
Db      22749 ACCTCCGAGGTTTTCGCGGCGTCGATGTCGGACTGCGCCGACGCATTGGCGCCACACCTG 22808

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              || |||||  ||  ||||   | ||||||    |  ||  ||  | ||||       ||  |  | |
Db      22809 GATTGGTCCCTGCTGGATGTGCTGCGCAACGCGGCCGGCGCTGCGCACCTTGACCACGAC 22868

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              |||||  ||||||||  | |||||||   | ||||||| ||   | ||  |   ||  ||||||||
Db      22869 GATGTCGTCCAGCCCGCGCTGTTCGCCATCATGGTCTCGCTCGCGGAGCTCTGGCGTTCG 22928

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              |  ||  |    ||  |  ||||||  | ||  ||  ||  |||||  ||||||||||  ||
```

```
Qy        418 GTGGAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGC 477
              | | ||| ||||   |  ||  | ||| | ||  || ||  ||| || ||||||||||||
Db       8682 GAGAAGCTGCTCGCCGACGACAGCTGGACCGGCAGGCTGGAGGTCGCCGCGGTCAACGGC 8741

Qy        478 CCCCGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGT 537
              |||  | ||   |  ||  | || || ||| |||||| |  |||| | |||| ||||||
Db       8742 CCCGACGCCGTGGTGGTGGCCGGTGACGCCCAGGCGGCGCGCGAGTTCCTGGAGTACTGC 8801

Qy        538 GCCGGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCAT 597
              |  |||   ||  | || |||||   ||||||||| ||||| |||||||||  | | ||
Db       8802 GAGGGCGTGGGCATCCGCGCCCGCGCGATCCCGGTGGACTACGCCTCGCACACCGCGCAC 8861

Qy        598 GTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCC 657
              ||  ||||| |||| || ||  || |  ||  |||||  |  |||| |||||  | | ||
Db       8862 GTCGAGCCCGTGCGCGACGAACTGGTCCAGGCGCTGGCCGGGATCACCCCGCGACGGGCC 8921

Qy        658 GGCGTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCC 717
              |  ||||||||||||||||  ||   |||  ||  || |||  |||     ||||||||
Db       8922 GAGGTGCCGTTCTTCTCCACCCTGACCGGCGACTTCCTCGACGGCACCGAGCTGGACGCG 8981

Qy        718 GCCTACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTG 777
              | ||||||||||||||||||  || |||||    |||  |   ||||| ||||| |||
Db       8982 GGCTACTGGTACCGCAACCTGCGTCACCCGGTGGAGTTCCACTCCGCCGTGCAGGCGCTG 9041

Qy        778 GCGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
                | ||  | ||| ||      ||| |||| |||||||| |||||    || |   |   |
Db       9042 ACCGACCAGGGATACGCGACGTTCATCGAGGTCAGCCCGCACCCGGTGCTGGCGTCGAGC 9101

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897
              || | || |||  ||| ||| |||     ||||   | | | |||| |  ||    |||
Db       9102 GTCCAGGAGACCCTCGACGACGCCGAGTCGGACGCGGCGGTGCTCGGGACGCTGGAACGC 9161

Qy        898 GGCGACAACGACACCCGCCGCTTC 921
              | ||    |||| ||  |||||||
Db       9162 GACGCGGGCGACGCCGACCGCTTC 9185


RESULT 8
US-08-439-009A-1
; Sequence 1, Application US/08439009A
; Patent No. 6004787
;  GENERAL INFORMATION:
;    APPLICANT:  Donadio, S
;    APPLICANT:  Katz, L
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Method of Directing Biosynthesis of
;    TITLE OF INVENTION:  Specific Polyketides
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steven F. Weinstock
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,009A
;      FILING DATE:  11-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Casuto, Dianne
;      REGISTRATION NUMBER:  40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847-938-3137
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11219 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 2338
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..6659
;      OTHER INFORMATION:  /function= "APPROXIMATE SPAN OF
;      OTHER INFORMATION:  MODULE 1"
;      OTHER INFORMATION:  /label= FUNCTION
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  744..11219
;      OTHER INFORMATION:  /function= "gene= "eryA""
;      OTHER INFORMATION:  /product= ""ORF1 encoding modules 1 & 2 for
;      OTHER INFORMATION:  6-deoxyerythronolide B""
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..1868
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1998..2198
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2250..3626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACP synthase domain/module1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3831..4811
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase 2 domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5574..6125
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6369..6626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 2 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..11219
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..8066
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACPsynthase of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  8262..9305
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9906..10454
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 2"
```

```
US-07-642-734C-1
; Sequence 1, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, L
;    APPLICANT:  Donadio, S
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Recombinant DNA Method for Producing
;    TITLE OF INVENTION:  Erythromycin Analogs
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Danckers, Andreas M
;      REGISTRATION NUMBER:  32652
;      REFERENCE/DOCKET NUMBER:  4952.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11219 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 2338
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..6659
;      OTHER INFORMATION:  /function= "APPROXIMATE SPAN OF
;      OTHER INFORMATION:  MODULE 1"
;      OTHER INFORMATION:  /label= FUNCTION
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  744..11219
;      OTHER INFORMATION:  /function= "gene= "eryA"."
;      OTHER INFORMATION:  /product= ""ORF1 encoding modules 1 & 2 for
;      OTHER INFORMATION:  6-deoxyerythronolide B""
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..1868
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1998..2198
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2250..3626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACP synthase domain/module1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3831..4811
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase 2 domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5574..6125
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6369..6626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 2 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..11219
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..8066
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACPsynthase of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  8262..9305
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9906..10454
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10707..10964
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 2"
US-07-642-734C-1

  Query Match             38.0%;  Score 349.6;  DB 1;  Length 11219;
  Best Local Similarity   62.7%;  Pred. No. 1.6e-52;
  Matches  579;  Conservative    0;  Mismatches  339;  Indels    6;  Gaps    2;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              || ||||| || || || ||||||| |||||||  || |||   | |||   | |
Db       8265 GTGTTCGTCTTCCCTGGGCAGGGTGCCCAGTGGGAAGGCATG---GCGCGGGAGTTGTTG 8321

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
                |  ||| || ||||||  |  |||    | ||||||  | ||   || ||  |  ||
Db       8322 CCGGTTCCCGTCTTCGCCGAGTCGATCGCCGAGTGCGATGCGGTGTTGTCGGAGGTGGCC 8381

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              |  |  || |||   ||  | |||   |       |||  |   ||  |  |||||||| |
Db       8382 GGATTCTCGGTGTCCGAGGTGCTGGAGCCACGTCCGGACGCGCCGTCGCTGGAGCGGGTC 8441

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              || ||||| ||||| |||||||||   || ||||||||  ||||     ||||||| ||
Db       8442 GACGTGGTGCAGCCGGTGCTGTTCGCGGTGATGGTGTCGCTGGCGCGGTTGTGGCGTGCC 8501

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              | ||||  ||  ||    || ||| | || || || ||||| |||||||| ||||||
Db       8502 TGCGGTGCCGTTCCTTCGGCCGTCATAGGGCACTCGCAGGGTGAGATCGCCGCCGCGGTG 8561

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              |||  || ||| ||   ||| ||||||      || || |  ||     |||||
Db       8562 GTGGCGGGAGCGTTGTCGCTGGAGGACGGCATGCGCGTCGTCGCCCGCCGGTCGAGGGCG 8621

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGC---CGCTGCCTGCCCAGGAG 417
              ||  || ||  |||||||| |||||  ||| ||||        |   || | |
Db       8622 GTGCGTGCGGTCGCGGGCCGGGGGAGCATGCTCTCGGTGCGCGGCGGCCGCTCCGACGTC 8681
```

```
Db        550 GGTGAGGGGGTGCACGCCCGGCCGATCCCCGGCGTCGACACGGCGGGCCACTCGCCGCAG 609

Qy        598 GTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCC 657

Db        610 GTGGACGCGTTGCGGGCTCATCTGCTGGAGGTGCTGGCCCCGGTCGCCCCCCGACCGGCC 669

Qy        658 GGCGTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCC 717

Db        670 GACATCCCGTTCTACTCGACGGTGACCGGCGGGCTGCTGGACGGCACCGAGCTGGACGCG 729

Qy        718 GCCTACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTG 777

Db        730 ACGTACTGGTACCGCAACATGCGCGAGCCCGTCGAGTTCGAGCGGGCCACACGGGCGCTG 789

Qy        778 GCGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837

Db        790 ATCGCCGACGGGCACGACGTCTTCCTGGAGACGAGCCCGCATCCCATGCTGGCCGTGGCG 849

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897

Db        850 CTGGAGCAGACGGTCACCGACGCCGGCACCGACGCGGCGGTGCTCGGGACCCTGCGCCGC 909

Qy        898 GGCGACAACGACACCCGC 915

Db        910 CGCCACGGCGGTCCTCGC 927


RESULT 6
US-08-997-467-29
; Sequence 29, Application US/08997467
; Patent No. 6200813
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/997,467
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/858,003
;      FILING DATE:  16-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1010 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-997-467-29

  Query Match             39.4%;  Score 362.8;  DB 4;  Length 1010;
  Best Local Similarity   63.7%;  Pred. No. 9.3e-55;
  Matches  585;  Conservative    0;  Mismatches  327;  Indels    6;  Gaps    2;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60

Db         13 GTGTTCGTGTTCCCCGGCCAGGGCTCGCAGTGGGCCGGAATGGCCGAGGGGCTGCTGGAG 72

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120

Db         73 CGGTCCGGCGCGTTCCGGAGTGCGGCCGACTCGTGCGACGCCGCGCTGCGGCCGTACCTC 132

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180

Db        133 GGCTGGTCGGTGCTGAGCGTGCTGCGCGGGGAACCGGACGCGCCCTCGCTCGACCGGGTC 192

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240

Db        193 GACGTCGTGCAGCCGGTGCTGTTCACGATGATGGTCTCGCTCGCGGCGGTCTGGCGTGCG 252

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300

Db        253 CTGGGGGTGGAACCGGCGGCGGTCGTCGGGCACTCGCAGGGTGAGATCGCCGCTGCCCAT 312

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360

Db        313 GTCGCCGGTGCGCTGTCGCTGGACGACTCGGCCCGGATCGTCGCCCTGCGCAGTCGGGCG 372

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420

Db        373 TGGCTCGGACTGGCGGGCAAGGGCGGCATGGTGGCGGTGCCGATGCCGGCGGAGGAGCTG 432

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480

Db        433 CGGCCGCGGCT---GGTGACGTGGGGGGACCGTCTGGCCGTCGCCGCCGTCAACAGCCCC 489

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540

Db        490 GGTTCCTGCGCCGTCGCAGGCGACCCGGAGGCGCTGGCCGAACTGGTGGCGCTGCTGACC 549

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCC---GGTCGACTATGCCTCGCACTGCCCCCAT 597

Db        550 GGTGAGGGGGTGCACGCCCGGCCGATCCCCGGCGTCGACACGGCGGGCCACTCGCCGCAG 609

Qy        598 GTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCC 657

Db        610 GTGGACGCGTTGCGGGCTCATCTGCTGGAGGTGCTGGCCCCGGTCGCCCCCCGACCGGCC 669

Qy        658 GGCGTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCC 717

Db        670 GACATCCCGTTCTACTCGACGGTGACCGGCGGGCTGCTGGACGGCACCGAGCTGGACGCG 729

Qy        718 GCCTACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTG 777

Db        730 ACGTACTGGTACCGCAACATGCGCGAGCCCGTCGAGTTCGAGCGGGCCACACGGGCGCTG 789

Qy        778 GCGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837

Db        790 ATCGCCGACGGGCACGACGTCTTCCTGGAGACGAGCCCGCATCCCATGCTGGCCGTGGCG 849

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897

Db        850 CTGGAGCAGACGGTCACCGACGCCGGCACCGACGCGGCGGTGCTCGGGACCCTGCGCCGC 909

Qy        898 GGCGACAACGACACCCGC 915

Db        910 CGCCACGGCGGTCCTCGC 927


RESULT 7
```

```
Db        133 GGCTGGTCGGTGCTGAGCGTGCTGCGCGGGGAACCGGACGCGCCCTCGCTCGACCGGGTC 192

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240

Db        193 GACGTCGTGCAGCCGGTGCTGTTCACGATGATGGTCTCGCTCGCGGCGGTCTGGCGTGCG 252

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300

Db        253 CTGGGGGTGGAACCGGCGGCGGTCGTCGGGCACTCGCAGGGTGAGATCGCCGCTGCCCAT 312

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360

Db        313 GTCGCCGGTGCGCTGTCGCTGGACGACTCGGCCCGGATCGTCGCCCTGCGCAGTCGGGCG 372

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420

Db        373 TGGCTCGGACTGGCGGGCAAGGGCGGCATGGTGGCGGTGCCGATGCCGGCGGAGGAGCTG 432

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480

Db        433 CGGCCGCGGCT---GGTGACGTGGGGGGACCGTCTGGCCGTCGCCGCCGTCAACAGCCCC 489

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540

Db        490 GGTTCCTGCGCCGTCGCAGGCGACCCGGAGGCGCTGGCCGAACTGGTGGCGCTGCTGACC 549

Qy        541 GGCACCGGGGGTGCGGGCCCGGCGGATCCC---GGTCGACTATGCCTCGCACTGCCCCCAT 597

Db        550 GGTGAGGGGGTGCACGCCCGGCCGATCCCCGGCGTCGACACGGCGGGCCACTCGCCGCAG 609

Qy        598 GTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCC 657

Db        610 GTGGACGCGTTGCGGGCTCATCTGCTGGAGGTGCTGGCCCCGGTCGCCCCCCGACCGGCC 669

Qy        658 GGCGTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCC 717

Db        670 GACATCCCGTTCTACTCGACGGTGACCGGCGGGCTGCTGGACGGCACCGAGCTGGACGCG 729

Qy        718 GCCTACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTG 777

Db        730 ACGTACTGGTACCGCAACATGCGCGAGCCCGTCGAGTTCGAGCGGGCCACACGGGCGCTG 789

Qy        778 GCGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837

Db        790 ATCGCCGACGGGCACGACGTCTTCCTGGAGACGAGCCCGCATCCCATGCTGGCCGTGGCG 849

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897

Db        850 CTGGAGCAGACGGTCACCGACGCCGGCACCGACGCGGCGGTGCTCGGGACCCTGCGCCGC 909

Qy        898 GGCGACAACGACACCCGC 915

Db        910 CGCCACGGCGGTCCTCGC 927


RESULT 5
US-09-078-166-29
; Sequence 29, Application US/09078166
; Patent No. 6063561
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  44
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/078,166
;      FILING DATE:  16-MAY-1979
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1010 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-09-078-166-29

  Query Match           39.4%;  Score 362.8;  DB 3;  Length 1010;
  Best Local Similarity  63.7%;  Pred. No. 9.3e-55;
  Matches  585;  Conservative    0;  Mismatches  327;  Indels    6;  Gaps    2;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60

Db         13 GTGTTCGTGTTCCCCGGCCAGGGCTCGCAGTGGGCCGGAATGGCCGAGGGGCTGCTGGAG 72

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120

Db         73 CGGTCCGGCGCGTTCCGGAGTGCGGCCGACTCGTGCGACGCCGCGCTGCGGCCGTACCTC 132

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180

Db        133 GGCTGGTCGGTGCTGAGCGTGCTGCGCGGGGAACCGGACGCGCCCTCGCTCGACCGGGTC 192

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240

Db        193 GACGTCGTGCAGCCGGTGCTGTTCACGATGATGGTCTCGCTCGCGGCGGTCTGGCGTGCG 252

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300

Db        253 CTGGGGGTGGAACCGGCGGCGGTCGTCGGGCACTCGCAGGGTGAGATCGCCGCTGCCCAT 312

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360

Db        313 GTCGCCGGTGCGCTGTCGCTGGACGACTCGGCCCGGATCGTCGCCCTGCGCAGTCGGGCG 372

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420

Db        373 TGGCTCGGACTGGCGGGCAAGGGCGGCATGGTGGCGGTGCCGATGCCGGCGGAGGAGCTG 432

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480

Db        433 CGGCCGCGGCT---GGTGACGTGGGGGGACCGTCTGGCCGTCGCCGCCGTCAACAGCCCC 489

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540

Db        490 GGTTCCTGCGCCGTCGCAGGCGACCCGGAGGCGCTGGCCGAACTGGTGGCGCTGCTGACC 549

Qy        541 GGCACCGGGGGTGCGGGCCCGGCGGATCCC---GGTCGACTATGCCTCGCACTGCCCCCAT 597
```

```
;  EARLIER APPLICATION NUMBER: 60/120,254
;  EARLIER FILING DATE: 1999-02-16
;  EARLIER APPLICATION NUMBER: 60/106,100
;  EARLIER FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 50937
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artifical Sequence: Recombinant DNA
US-09-428-517-1

  Query Match             40.6%;  Score 373.8;  DB 4;  Length 50937;
  Best Local Similarity   64.4%;  Pred. No. 1.1e-56;
  Matches  593;  Conservative    0;  Mismatches  322;  Indels    6;  Gaps    2;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              || ||||| ||||| || ||||||  ||||||||  |||||||  |   |  |||| |
Db      24645 GTGTTCGTCTTTCCTGGTCAGGGTTCGCAGTGGGTGGGGATGGCGGCTGGGTTGCTGGAT 24704

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
               | | |||||||||||||    | | ||  || ||| |  | ||  ||||  |||  |  |
Db      24705 GCGTGTCCGGTGTTCGCGGAGGCGGTGGCGGAGTGTGCCGCGGTGCTGGATCCGGTGACG 24764

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              |  |||||  |||| ||  |  |||||   |  |  || || || ||     | |||||
Db      24765 GGTTGGTCGCTGGTCGAGGTGTTGCGC---GGTGGTGGTGAGGCTGTTCTTGGGCGGGTT 24821

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              |||||||| ||||| | | |||      || ||||||||  ||||     |||||  |
Db      24822 GATGTGGTGCAGCCGGCGTTGTGGGCGGTGATGGTGTCACTGGCCCGGACCTGGCGGTAT 24881

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              |||||| | || || |  |||||  | || |||||  ||||| ||||||||  || ||   |
Db      24882 TACGGTGTGGAGCCTGCTGCGGTTGTGGGGCATTCGCAGGGTGAGATCGCTGCGGCTTGT 24941

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              |||  ||||| | |||| |||   || |  |||  |   || | | |||| ||||||||
Db      24942 GTGGCTGGGGGGTTGAGTCTGGCCGATGGTGCGCGGGTGGTGGTGTTGCGGAGCCGGGCG 25001

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              | |||    |   ||  |||||||||||| ||| ||   ||  | ||    |||
Db      25002 ATCGCCCGGATCGCTGGTGGGGGCGGCATGGTCTCGGTCGGTCTTTCAGC---TGAGCGT 25058

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              |  |    |||    ||    |  |  || ||| | | ||| ||||||||||| |||||
Db      25059 GTCCGCACCATGCTCGACACCTACGGTGGCCGGGTTTCGGTCGCGGCGGTCAATGGCCCG 25118

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              |||  |||| ||| || || || | | ||||  |||| ||| || ||||    ||||
Db      25119 TCCTCGACCGTCGTGTCCGGTGACGTCCAGGCCCTGGATGAGTTGTTGGCCGGTTGTGAG 25178

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              |    || || ||||| || ||  ||||||| |||||||||||| |||| | | ||  ||
Db      25179 CGGGAGGGTGTCCGGGCTCGTCGTGTCCCGGTGGACTATGCCTCCCACTCCGCGCAGATG 25238

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              | |   | || || ||| |||||||    ||||| ||||||||  ||||| |  ||| |
Db      25239 GACCAGTTACGCGATGAGCTGCTGGAAGCGCTGGCGGACATCACTCCGCAACATTCCAGT 25298

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              || |||||||||| ||||||  ||     |||||||||||| ||  | ||||| || |
Db      25299 GTTCCGTTCTTCTCGACGGTGACGGCGGACTGGCTGGACACGACCGCTCTGGATGCGGGG 25358

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              |||||||  |  || |||    || |||||||  |||   || |||||| | |  || | |
Db      25359 TACTGGTTCACGAATCTGCGGGAGACGGTCCGGTTCCAGGAAGCCGTCGAAGGGCTCGTG 25418

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              | | | ||     |||  ||||||||   |||||| ||||||  ||||||||||  |||
Db      25419 GCTCAGGGGATGGGCGCGTTCGTCGAGTGCAGCCCGCACCCCGTCCTCGTCCCCGGTATC 25478

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
              ||  | |||  ||| | |  || | |  ||| | |||    ||||| ||| || || | |
Db      25479 GAGCAGACCCTCGACGCCCTCGACCAGAACGCCGCCGTACTCGGCTCCCTGCGGCGTGAC 25538

Qy        901 GACAACGACACCCGCCGCTTC 921
              ||   || |     |||  ||
Db      25539 GAAGGCGGCCTGGACCGACTC 25559


RESULT 4
US-08-858-003-29
; Sequence 29, Application US/08858003
; Patent No. 6060234
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, Leonard
;    APPLICANT:  Stassi, Diane L.
;    APPLICANT:  Summers Jr., Richard G.
;    APPLICANT:  Ruan, Xiaoan
;    APPLICANT:  Pereda-Lopez, Ana
;    APPLICANT:  Kakavas, Stephan J.
;    TITLE OF INVENTION:  NOVEL POLYKETIDE DERIVATIVES
;    TITLE OF INVENTION:  AND RECOMBINANT METHODS FOR MAKING SAME
;    NUMBER OF SEQUENCES:  34
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Rd.
;      CITY:  Abbott Park
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/858,003
;      FILING DATE:  16-MAY-1979
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dianne Casuto
;      REGISTRATION NUMBER:  P-40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (847)-938-3137
;      TELEFAX:  (847)-938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  29:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1010 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-858-003-29

  Query Match             39.4%;  Score 362.8;  DB 3;  Length 1010;
  Best Local Similarity   63.7%;  Pred. No. 9.3e-55;
  Matches  585;  Conservative    0;  Mismatches  327;  Indels    6;  Gaps    2;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              || |||||||| ||||| |||||    |||||||||||| ||||  | | | ||||| |
Db         13 GTGTTCGTGTTCCCCGGCCAGGGCTCGCAGTGGGCCGGAATGGCCGAGGGGCTGCTGGAG 72

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
               ||    | ||||    |   |    |  ||||||| |  || |||  |||||   | .
Db         73 CGGTCCGGCGCGTTCCGGAGTGCGGCCGACTCGTGCGACGCCGCGCTGCGGCCGTACCTC 132

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              | |||||| ||| ||  | | |||||| ||||  |||  |   |   |   || |||| |
```

```
;       NAME/KEY:  misc_feature
;       LOCATION:  10225..10483
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 4"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  10723..20235
;       OTHER INFORMATION:  /codon_start= 10723
;       OTHER INFORMATION:  /function= "gene =eryA"
;       OTHER INFORMATION:  /product= "orf3 encoding modules 5 & 6
;       OTHER INFORMATION:  6-deoxyerythronolide B formatio"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10723..15165
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10831..12174
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  12379..13350
;       OTHER INFORMATION:  /function= "approximatr span of
;       OTHER INFORMATION:  acyltransferase domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14062..14610
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14857..15114
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15166..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15172..16569
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  16768..17721
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  18379..18921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19149..19398
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19492..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  thioesterase domain of module 6"
US-08-439-009A-3

  Query Match             41.8%;  Score 385.4;  DB 3;  Length 20235;
  Best Local Similarity   65.6%;  Pred. No. 1.1e-58;
  Matches  579;  Conservative    0;  Mismatches  301;  Indels    3;  Gaps    1;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              |||||| |||| || || ||||||  ||| |||||||| |||||||     |||||  |
Db       1696 GTCTTCCTGTTCCCGGGGCAGGGTTCGCAATGGGCCGGCATGGGTGCCGAGCTGCTGTCG 1755

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              || || |||||||||||||| ||  ||| | ||| ||||| ||| |  |||| ||| |  |
Db       1756 TCGTCGCCGGTGTTCGCCGGGAAGATCCGGGCCTGCGACGAGTCGATGGCCCCGATGCAG 1815

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||   ||    ||| | ||||| | ||   ||||||   |||     ||| |||| |
Db       1816 GACTGGAAGGTCTCCGACGTGCTGCGTCAGGCGCCGGGGGCGCCGGGCCTGGACCGGGTC 1875

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              || ||||| ||||| ||| ||||||   || ||||||||  |||| |  |||||||| ||
Db       1876 GACGTGGTGCAGCCGGTGTTGTTCGCGGTGATGGTGTCGCTGGCGGAGCTGTGGCGCTCG 1935

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              |||||  | || ||||  |||||| | || || || ||||||||||||||| ||||||||
Db       1936 TACGGCGTGGAGCCCGCGGCGGTCGTGGGGCACTCGCAGGGCGAGATCGCCGCCGCGCAC 1995

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||    |||||||| |   || ||||||||||||||   || | |   ||||||||    |
Db       1996 GTCGCCGGGGCGCTCACGTTGGAGGACGCGGCGAAGCTCGTCGTGGGCCGCAGCCGCCTG 2055

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              ||    |  |   ||  |||||||||||||| | ||  |||||   |     | ||||
Db       2056 ATGCGGTCGCTCTCCGGGGAGGGCGGCATGGCCGCCGTCGCGCTGGGCGAGGCCGCGGTG 2115

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
                 |||| | |   || ||||  || ||| | | |||||| ||||||||||| |||
Db       2116 CGCGAGCGCCTGCG---GCCGTGGCAGGACCGGCTCTCGGTGGCCGCGGTCAACGGTCCC 2172

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              || ||   ||  ||||| || ||  |||  ||| ||   | | |    | | |||| ||
Db       2173 CGGTCGGTCGTGGTCTCCGGCGAGCCCGGCGCGCTGCGGGCGTTTTCCGAGGACTGCGCG 2232

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              | |   ||  | || | |||    |||   || ||||| ||||||||||  ||.||  |
Db       2233 GCCGAGGGCATCCGCGTCCGCGACATCGACGTGGACTACGCCTCGCACTCGCCGCAGATC 2292

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              |||   | || |||||  | || ||    |   || ||||||   ||||  |||  |  |
Db       2293 GAGCGGGTCCGCGAGGAACTCCTCGAAACGACCGGCGACATCGCGCCGCGCCCGGCGCGG 2352

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              ||| |||||  |||||| ||||||       |   |||||  |||     ||||| |||
Db       2353 GTGACGTTCCACTCCACTGTGGAGTCGCGGTCTATGGACGGCACCGAGCTGGATGCCCGG 2412

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              |||||||||||||||||| | || |||| || |||  |||  |||||   |   ||||||
Db       2413 TACTGGTACCGCAACCTGCGCGAGACGGTGCGCTTCGCCGACGCCGTGACGCGGCTGGCG 2472

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              ||    |||  ||  ||  ||| |||| |||||||| ||  ||    | | ||||  ||| ||
Db       2473 GAGTCGGGATACGACGCGTTCATCGAGGTCAGCCCGCATCCGGTCGTGGTCCAGGCCGTC 2532

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCG 883
              || ||  |   ||||||  | | ||  | ||    |||| |||
Db       2533 GAGGAGGCGGTCGAAGAGGCTGACGGTGCCGAAGACGCGGTCG 2575


RESULT 3
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Shah, Sanjay Krishnakant
;  APPLICANT: McDaniel, Robert
;  APPLICANT: Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/428,517
;  CURRENT FILING DATE: 1999-10-28
```

```
Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||     |||||||| |   || ||||||||||||||    || | |    ||||||||    |
Db       1996 GTCGCCGGGGCGCTCACGTTGGAGGACGCGGCGAAGCTCGTCGTGGGCCGCAGCCGCCTG 2055

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
               ||    |  |    ||   ||||||||||||| | ||  |||||  |      | ||||
Db       2056 ATGCGGTCGCTCTCCGGGGAGGGCGGCATGGCCGCCGTCGCGCTGGGCGAGGCCGCGGTG 2115

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              ||| | |  |   || ||||  ||  ||| | | |||||| ||||||||||| |||
Db       2116 CGCGAGCGCCTGCG---GCCGTGGCAGGACCGGCTCTCGGTGGCCGCGGTCAACGGTCCC 2172

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              || ||   ||  ||||| || ||  |||  ||| ||   | | |     | | |||| ||
Db       2173 CGGTCGGTCGTGGTCTCCGGCGAGCCCGGCGCGCTGCGGGCGTTTTCCGAGGACTGCGCG 2232

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              | |   ||  | || | |||     |||   || ||||| ||||||||||  || ||  |
Db       2233 GCCGAGGGCATCCGCGTCCGCGACATCGACGTGGACTACGCCTCGCACTCGCCGCAGATC 2292

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
               |||   | || |||||  | || ||   |   || ||||||   ||||  ||| |  |
Db       2293 GAGCGGGTCCGCGAGGAACTCCTCGAAACGACCGGCGACATCGCGCCGCGCCCGGCGCGG 2352

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              ||| |||||  ||||||  ||||||       |  |||||  |||    ||||| |||
Db       2353 GTGACGTTCCACTCCACTGTGGAGTCGCGGTCTATGGACGGCACCGAGCTGGATGCCCGG 2412

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              ||||||||||||||||||| | || |||| || |||  ||| |||||    |   ||||||
Db       2413 TACTGGTACCGCAACCTGCGCGAGACGGTGCGCTTCGCCGACGCCGTGACGCGGCTGGCG 2472

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              ||    ||| ||  ||  ||| |||| |||||||| || ||   | | ||||  ||| ||
Db       2473 GAGTCGGGATACGACGCGTTCATCGAGGTCAGCCCGCATCCGGTCGTGGTCCAGGCCGTC 2532

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCG 883
              || ||  |   ||||||  | | ||  | ||    |||| |||
Db       2533 GAGGAGGCGGTCGAAGAGGCTGACGGTGCCGAAGACGCGGTCG 2575


RESULT 2
US-08-439-009A-3
; Sequence 3, Application US/08439009A
; Patent No. 6004787
;  GENERAL INFORMATION:
;    APPLICANT:  Donadio, S
;    APPLICANT:  Katz, L
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Method of Directing Biosynthesis of
;    TITLE OF INVENTION:  Specific Polyketides
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steven F. Weinstock
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,009A
;      FILING DATE:  11-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Casuto, Dianne
;      REGISTRATION NUMBER:  40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847-938-3137
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20235 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 238
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  19..10722
;      OTHER INFORMATION:  /codon_start= 19
;      OTHER INFORMATION:  /function= "gene eryA"
;      OTHER INFORMATION:  /product= "eryA ORF2 encoding modules 3 & 4 for
;      OTHER INFORMATION:  6-deoxyerythronolide B"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19..4470
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  97..1482
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACP synthase of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1693..2670
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3406..3921
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4171..4428
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4471..10722
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 4"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4471..5847
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACPsynhase domain of module"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6054..7026
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 4"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  7165..9216
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  dehydratase and enoylreductase domains m"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9433..9984
;      OTHER INFORMATION:  /function= "approximate span
;      OTHER INFORMATION:  beta-ketoreductase of module 4"
;    FEATURE:
```

```
;       OTHER INFORMATION:  /codon_start= 19
;       OTHER INFORMATION:  /function= "gene eryA"
;       OTHER INFORMATION:  /product= "eryA ORF2 encoding modules 3 & 4 for
;       OTHER INFORMATION:  6-deoxyerythronolide B"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19..4470
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  97..1482
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacyl ACP synthase of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1693..2670
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3406..3921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4171..4428
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4471..10722
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4471..5847
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynhase domain of module"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6054..7026
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  7165..9216
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  dehydratase and enoylreductase domains m"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  9433..9984
;       OTHER INFORMATION:  /function= "approximate span
;       OTHER INFORMATION:  beta-ketoreductase of module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10225..10483
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 4"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  10723..20235
;       OTHER INFORMATION:  /codon_start= 10723
;       OTHER INFORMATION:  /function= "gene =eryA"
;       OTHER INFORMATION:  /product= "orf3 encoding modules 5 & 6
;       OTHER INFORMATION:  6-deoxyerythronolide B formatio"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10723..15165
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10831..12174
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  12379..13350
;       OTHER INFORMATION:  /function= "approximatr span of
;       OTHER INFORMATION:  acyltransferase domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14062..14610
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14857..15114
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15166..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15172..16569
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  16768..17721
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  18379..18921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19149..19398
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19492..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  thioesterase domain of module 6"
US-07-642-734C-3

  Query Match             41.8%;  Score 385.4;  DB 1;  Length 20235;
  Best Local Similarity   65.6%;  Pred. No. 1.1e-58;
  Matches  579;  Conservative    0;  Mismatches  301;  Indels    3;  Gaps    1;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              |||||| |||| || || |||||  ||| ||||||||  |||||||     |||||  |
Db       1696 GTCTTCCTGTTCCCGGGGCAGGGTTCGCAATGGGCCGGCATGGGTGCCGAGCTGCTGTCG 1755

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              || || |||||||||||| ||  ||| | ||| ||||| ||| |  |||| |||  |  |
Db       1756 TCGTCGCCGGTGTTCGCCGGGAAGATCCGGGCCTGCGACGAGTCGATGGCCCCGATGCAG 1815

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||   ||    ||| | ||||| | ||   ||||||   |||     ||| |||| |
Db       1816 GACTGGAAGGTCTCCGACGTGCTGCGTCAGGCGCCGGGGGCGCCGGGCCTGGACCGGGTC 1875

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              || ||||| ||||| ||| |||||   || ||||||||  |||| |  |||||||| ||
Db       1876 GACGTGGTGCAGCCGGTGTTGTTCGCGGTGATGGTGTCGCTGGCGGAGCTGTGGCGCTCG 1935

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              |||||  | || ||||  |||||| | || || || ||||||||||||||| ||||||||
Db       1936 TACGGCGTGGAGCCCGCGGCGGTCGTGGGGCACTCGCAGGGCGAGATCGCCGCCGCGCAC 1995
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 39.8967 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_3148_4068
Perfect score: 921
Sequence: 1 gtcttcgtgtttcccggaca.....acaacgacacccgccgcttc 921

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	385.4	41.8	20235	1	US-07-642-734C-3	Sequence 3, Appli
2	385.4	41.8	20235	3	US-08-439-009A-3	Sequence 3, Appli
3	373.8	40.6	50937	4	US-09-428-517-1	Sequence 1, Appli
4	362.8	39.4	1010	3	US-08-858-003-29	Sequence 29, Appl
5	362.8	39.4	1010	3	US-09-078-166-29	Sequence 29, Appl
6	362.8	39.4	1010	4	US-08-997-467-29	Sequence 29, Appl
7	349.6	38.0	11219	1	US-07-642-734C-1	Sequence 1, Appli
8	349.6	38.0	11219	3	US-08-439-009A-1	Sequence 1, Appli
9	348	37.8	80161	3	US-09-036-987A-1	Sequence 1, Appli
10	348	37.8	80161	4	US-09-370-700-1	Sequence 1, Appli
11	345	37.5	28958	1	US-08-258-261B-6	Sequence 6, Appli
12	345	37.5	28958	1	US-08-456-837-6	Sequence 6, Appli
13	345	37.5	28958	1	US-08-457-342-6	Sequence 6, Appli
14	345	37.5	28958	1	US-08-457-646A-6	Sequence 6, Appli
15	345	37.5	28958	1	US-08-458-076A-6	Sequence 6, Appli
16	345	37.5	28958	1	US-08-764-233A-4	Sequence 4, Appli
17	345	37.5	28958	1	US-08-457-335A-6	Sequence 6, Appli
18	345	37.5	28958	1	US-08-729-214-6	Sequence 6, Appli
19	345	37.5	28958	3	US-09-028-934-6	Sequence 6, Appli
20	345	37.5	49377	1	US-08-764-233A-1	Sequence 1, Appli
21	344.4	37.4	44377	2	US-08-804-227C-7	Sequence 7, Appli
22	344.4	37.4	44377	2	US-08-804-198-1	Sequence 1, Appli
23	342.2	37.2	43280	2	US-08-804-227C-1	Sequence 1, Appli
24	319	34.6	441529	4	US-09-103-840A-1	Sequence 1, Appli
25	318	34.5	13842	4	US-09-105-537-30	Sequence 30, Appl
26	318	34.5	36778	4	US-09-105-537-5	Sequence 5, Appli
27	318	34.5	38506	3	US-09-320-878-19	Sequence 19, Appl
28	311	33.8	11220	4	US-09-105-537-32	Sequence 32, Appl
29	308.6	33.5	1035	3	US-08-858-003-30	Sequence 30, Appl
30	308.6	33.5	1035	3	US-09-078-166-30	Sequence 30, Appl
31	308.6	33.5	1035	4	US-08-997-467-30	Sequence 30, Appl
32	308.6	33.5	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c 33	306.8	33.3	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c 34	306.8	33.3	4411529	4	US-09-103-840A-1	Sequence 1, Appli
35	305.8	33.2	4689	4	US-09-105-537-34	Sequence 34, Appl
c 36	302	32.8	1881	4	US-09-434-288-5	Sequence 5, Appli
37	298.4	32.4	68750	3	US-09-335-409-1	Sequence 1, Appli
38	298.4	32.4	68750	4	US-09-568-102-1	Sequence 1, Appli
39	298.4	32.4	68750	4	US-09-567-969-1	Sequence 1, Appli
40	298.4	32.4	68750	4	US-09-568-480-1	Sequence 1, Appli
41	298.4	32.4	68750	4	US-09-568-486-1	Sequence 1, Appli
42	298.4	32.4	68750	4	US-09-568-472-1	Sequence 1, Appli
43	298.4	32.4	68750	4	US-09-567-899-1	Sequence 1, Appli
44	296.8	32.2	71989	4	US-09-443-501A-2	Sequence 2, Appli
45	293.8	31.9	13987	2	US-08-804-227C-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-07-642-734C-3
; Sequence 3, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
; APPLICANT: Katz, L
; APPLICANT: Donadio, S
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Recombinant DNA Method for Producing
; TITLE OF INVENTION: Erythromycin Analogs
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edward H. Gorman
; STREET: Abbott Laboratories D377/AP6D-2 One Abbott
; STREET: Park Rd
; CITY: Abbott Park
; STATE: IL
; COUNTRY: US
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/642,734C
; FILING DATE: 17-JAN-91
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Danckers, Andreas M
; REGISTRATION NUMBER: 32652
; REFERENCE/DOCKET NUMBER: 4952.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-9396
; TELEFAX: 708-938-2623
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20235 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Saccharopolyspora erythraea
; STRAIN: NRRL 238
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 19..10722

```
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match           47.7%;  Score 621.4;  DB 4;  Length 4411529;
  Best Local Similarity  68.7%;  Pred. No. 4.1e-113;
  Matches  894;  Conservative    0;  Mismatches  386;  Indels   21;  Gaps    2;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || ||| | ||   |  | || |||| |||||| |  ||||| || ||     || ||| ||
Db    1875404 GAGCCGGTGGCGGTGGTGGGGATGGGATGCCGCTATCCGGGTGGGGTGGATTCGCCGGAG 1875463

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
             |||||||||||||  ||    ||| | |||||  |   |  ||||||  | || |||
Db    1875464 ACGCTGTGGGAGCTGGTGGCACACGGCCGTGACGCGGTTTCGGAGTTCCCGGCGGATCGC 1875523

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||      ||     |      | ||  || |||      |   |||| ||  ||| |
Db    1875524 GGCTGGGATGTGGCGGGGTTGTTTGACCCCGATCCCGACGCGGTAGGCAAGTCGTATACC 1875583

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              | | | ||||||||| |    || |  ||    || |||||||  || ||||| |||
Db    1875584 CGGTGCGGCGGGTTCTTGACGGATGTCGCCGGTTTTGACGCCGAGTTTTTCGGGATCGCA 1875643

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              ||  | |||||  || ||||||||||| ||||||||| |||||  ||||   ||| |||||
Db    1875644 CCCAGCGAGGCGCTTGCGATGGATCCCCAGCAGCGGTTGCTGTTGGAAGTGTCGTGGGAA 1875703

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||||| |||||||||||| ||||| ||   |  || | || || || |  |||||||| |||
Db    1875704 GCGTTGGAGCGGGCGGGCATCGACCCAATCACGTTGCGGGGTTCGCAGACGGGCGTGTTC 1875763

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              || || |   ||  |  |  || ||                     |      ||    |
Db    1875764 GCCGGGGTGTTCCACGGCTCGTATGG--------------GGGCCAAGGCCGGGTGCCG 1875808

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              |  ||| ||||| || ||   ||    ||  | |||   |||||    || || ||  |
Db    1875809 GGTGACCTGGAGCGCTACGGGCTGCGTGGCTCGACGCTGAGCGTGGCCTCCGGGCGGGTG 1875868

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              || ||    || || ||| | || |||||| |   | ||||| || | ||| ||| | |||
Db    1875869 GCGTATGTGTTGGGCCTGCAGGGCCCGGCGGTGTCGGTGGATACCGCGTGTTCGTCGTCG 1875928

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              | ||| | |||||| |||||   ||||| |||||    || || |||   || ||||||
Db    1875929 TTGGTGGCACTGCATTTGGCGGTGCAGTCACTGCGCCTCGGCGAATGCGACCTGGCGCTG 1875988

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              | ||| || ||| | || ||| |||||| |  ||||||||||||||  | |||||||
Db    1875989 GTCGGTGGGGTCACCGTGATGGCCACCCCGGCGATGTTCATCGAGTTCAGCAGGCAGCGG 1876048

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              | |||||| |  || ||  | || ||||| ||  ||| ||| ||||| || |||| |
Db    1876049 GCGCTGTCCGCCGATGGTCGTTGTAAGGCCTATGCGGGTGCCGCCGATGGGACCGCGTTT 1876108

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              | |||||||| |||| ||||  || ||| ||||||| | || |||   ||| |||||||
Db    1876109 GCCGAGGGCGCCGGGGTGCTCGTGCTGGCGCGGTTGGCTGACGCGCGCCGGTTGGGGCAT 1876168

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              | |||||||||| |||| |||||    |||||||| |||||||| || || || ||||||
Db    1876169 CCGGTGCTGGCGCTGGTGCGCGGATCGGCGGTCAATCAGGACGGCGCCTCCAACGGGCTG 1876228

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              |  ||||||| || ||||| ||| | |||||||||      |||| |||| |   ||| ||
Db    1876229 GCCACGCCGAATGGGCCGGCGCAGCAACGGGTGATCACTGCGGCGCTGGCCAGTGCGCGG 1876288

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||    || || || |||||||||| |||||||||||||||||||| ||| |||| |||
Db    1876289 TTAGGTGTCGCCGACGTGGATGTGGTCGAGGGGCACGGGACGGGCACCACGTTGGGGGAT 1876348

Qy         961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
               || || ||||| |||||| |  | || ||||| || |||||| | |   ||  ||||| ||
Db    1876349 CCCATTGAGGCGCAGGCGATTTTGGCGACGTATGGACAGCGGCCGGCCGATCGGCCGTTG 1876408

Qy        1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
               ||||||||||  | || || |||||||| || ||   ||| || ||||| || |  |||
Db    1876409 TGGCTGGGGTCGATCAAATCGAACATCGGTCATACGTCGGCGGCTGCGGGGGTCGCCGGG 1876468

Qy        1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
               || |||||||||||  |||| ||||  | || ||| ||||   ||||||||| ||||||
Db    1876469 GTGATCAAGATGGTGCAGGCGATGCGCCACGGCGTGCTGCCCAAGACGTTGCACGTGGAT 1876528

Qy        1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
                |||| ||||||| ||||| ||||| ||||||||||||  | || |||| ||       |
Db    1876529 GTGCCGACGCCGCATGTGGATTGGTCGGCGGGGGCGGTGTCGTTGTTGACCGA------G 1876582

Qy        1201 CCGTGGCCGGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
               ||| ||||| || ||| | | || |||  ||||||||| || ||||||||||||||||||
Db    1876583 CCGCGGCCGTGGCACGTGCCGGGCCGGCCGCGGCGGGCCGGTGTGTCGTCGTTCGGGATC 1876642

Qy        1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGC 1301
               ||||||| || || |||||||||| |||| ||||| |||||
Db    1876643 AGCGGCACCAACGCACATGTGATTCTGGAAGAGGCACCGGC 1876683

Search completed: June 18, 2003, 01:06:42
Job time : 78.4013 secs
```

```
Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257
             ||    |||||||  | |   ||  ||  || |||| || ||||  || || || |||||
Db       1297 GTCGACTGGCCGGAGAAGCAGGACGGCGGGCTGCGCCGCGCGGCTGTCTCCTCCTTCGGC 1356

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
             |||  |||| |||||  |||||  ||   |  ||||||||||  ||||||
Db       1357 ATCAGCGGGACGAACGCGCACGTCGTCCTGGAGGAGGCCCCGGCG 1401


RESULT 14
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             47.7%;  Score 621.4;  DB 4;  Length 4403765;
  Best Local Similarity   68.7%;  Pred. No. 4.1e-113;
  Matches  894;  Conservative    0;  Mismatches  386;  Indels   21;  Gaps    2;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || ||| | ||   |  | || |||| |||||| |  ||||| || ||     || |||| ||
Db    1866281 GAGCCGGTGGCGGTGGTGGGGATGGGATGCCGCTATCCGGGTGGGGTGGATTCGCCGGAG 1866340

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              |||||||||||||  ||     ||| | |||||  |    |  ||||||  | || |||
Db    1866341 ACGCTGTGGGAGCTGGTGGCACACGGCCGTGACGCGGTTTCGGAGTTCCCGGCGGATCGC 1866400

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||      ||     |     | ||  || |||     |   |||| ||  ||| |
Db    1866401 GGCTGGGATGTGGCGGGGTTGTTTGACCCCGATCCCGACGCGGTAGGCAAGTCGTATACC 1866460

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              | | | |||||||||  |     ||  |  ||     || |||||||  || |||||  |||
Db    1866461 CGGTGCGGCGGGTTCTTGACGGATGTCGCCGGTTTTGACGCCGAGTTTTTCGGGATCGCA 1866520

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              ||  |  |||||  ||  ||||||||||  |||||||||  |||||  ||||    |||  |||||
Db    1866521 CCCAGCGAGGCGCTTGCGATGGATCCCCAGCAGCGGTTGCTGTTGGAAGTGTCGTGGGAA 1866580

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              |||||  ||||||||||  |||||  ||   |  || | || || || |   |||||||  |||
Db    1866581 GCGTTGGAGCGGGCGGGCATCGACCCAATCACGTTGCGGGGTTCGCAGACGGGCGTGTTC 1866640

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              || || |   ||  |  |   || ||                 |          ||   |
Db    1866641 GCCGGGGTGTTCCACGGCTCGTATGG---------------GGGCCAAGGCCGGGTGCCG 1866685

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              |  ||| |||||  || ||    ||    ||  | |||   |||||    || || || |
Db    1866686 GGTGACCTGGAGCGCTACGGGCTGCGTGGCTCGACGCTGAGCGTGGCCTCCGGGCGGGTG 1866745

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
```

```
              || ||    || || ||| | || ||||||  |  | |||||  || |  ||| |||  | |||
Db    1866746 GCGTATGTGTTGGGCCTGCAGGGCCCGGCGGTGTCGGTGGATACCGCGTGTTCGTCGTCG 1866805

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              |  |||  | ||||||  |||||    |||||  |||||     ||  ||  |||    ||  ||||||
Db    1866806 TTGGTGGCACTGCATTTGGCGGTGCAGTCACTGCGCCTCGGCGAATGCGACCTGGCGCTG 1866865

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              | |||  ||  |||  |  ||  |||  ||||||    |  ||||||||||||||    |  |||||||
Db    1866866 GTCGGTGGGGTCACCGTGATGGCCACCCCGGCGATGTTCATCGAGTTCAGCAGGCAGCGG 1866925

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              | ||||||  |  ||  ||  |  ||||||  ||  |||  |||  |||||  ||  ||||  |
Db    1866926 GCGCTGTCCGCCGATGGTCGTTGTAAGGCCTATGCGGGTGCCGCCGATGGGACCGCGTTT 1866985

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              | ||||||||  ||||  ||||   ||  |||  |||||||  |  ||  |||    |||  ||||||||
Db    1866986 GCCGAGGGCGCCGGGGTGCTCGTGCTGGCGCGGTTGGCTGACGCGCGCCGGTTGGGGCAT 1867045

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              | ||||||||||  ||||  ||||     ||||||||  ||||||||  ||  ||  ||  ||||||
Db    1867046 CCGGTGCTGGCGCTGGTGCGCGGATCGGCGGTCAATCAGGACGGCGCCTCCAACGGGCTG 1867105

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              |  |||||||  ||  |||||  |||  |  |||||||||      ||||  ||||  |   |||  ||
Db    1867106 GCCACGCCGAATGGGCCGGCGCAGCAACGGGTGATCACTGCGGCGCTGGCCAGTGCGCGG 1867165

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||     ||  ||  ||  ||||||||||  ||||||||||||||||||||||||  |||  ||||  |||
Db    1867166 TTAGGTGTCGCCGACGTGGATGTGGTCGAGGGGCACGGGACGGGCACCACGTTGGGGGAT 1867225

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              || || |||||  ||||||  |  |  ||  |||||  ||  ||||||  |  |   ||  |||||  ||
Db    1867226 CCCATTGAGGCGCAGGCGATTTTGGCGACGTATGGACAGCGGCCGGCCGATCGGCCGTTG 1867285

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              ||||||||||  |  ||  ||  ||||||||  ||  ||    |||  ||  |||||  ||  |  |||
Db    1867286 TGGCTGGGGTCGATCAAATCGAACATCGGTCATACGTCGGCGGCTGCGGGGGTCGCCGGG 1867345

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              ||  ||||||||||||  ||||  ||||  |  ||  ||  ||||    |||||||||  ||||||
Db    1867346 GTGATCAAGATGGTGCAGGCGATGCGCCACGGCGTGCTGCCCAAGACGTTGCACGTGGAT 1867405

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              |||| |||||||  |||||  |||||  ||||||||||||  |  ||  ||||  ||       |
Db    1867406 GTGCCGACGCCGCATGTGGATTGGTCGGCGGGGGCGGTGTCGTTGTTGACCGA------G 1867459

Qy       1201 CCGTGGCCGGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              |||  |||||  ||  |||  |  |  ||  |||   |||||||||  ||  ||||||||||||||||||
Db    1867460 CCGCGGCCGTGGCACGTGCCGGGCCGGCCGCGGCGGGCCGGTGTGTCGTCGTTCGGGATC 1867519

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGC 1301
              |||||||  ||  ||  |||||||||  ||||  |||||  |||||
Db    1867520 AGCGGCACCAACGCACATGTGATTCTGGAAGAGGCACCGGC 1867560


RESULT 15
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
```

```
Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
Db        814 TCCGAGGGCGTCGGCGTCCTCCTCGTCGAGCGCCTGTCCGACGCGCGTCGCAAGGGCCAT 873

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
Db        874 CGGATCCTCGCCGTGGTCCGGGGCACCGCCGTCAACCAGGACGGCGCCAGCAGCGGCCTC 933

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
Db        934 ACGGCTCCGAACGGGCCGTCGCAGCAGCGCGTCATCCGACGTGCCCTGGCGGACGCCCGG 993

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
Db        994 CTCACGACCTCCGACGTGGACGTCGTCGAGGCCCACGGCACGGGTACGCGACTCGGCGAC 1053

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAG---CGGGCCGGTGACAGGCCG 1017
Db       1054 CCGATCGAGGCGCAGGCCGTCATCGCCACGTACGGGCAGGGCCGTGACGGCGAACAGCCG 1113

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077
Db       1114 CTGCGCCTCGGGTCGTTGAAGTCCAACATCGGACACACCCAGGCCGCCGCCGGTGTCTCC 1173

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137
Db       1174 GGCGTGATCAAGATGGTCCAGGCGATGCGCCACGGCGTCCTGCCGAAGACGCTCCACGTG 1233

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197
Db       1234 GAGAAGCCGACGGACCAGGTGGACTGGTCCGCGGGCGCGGTCGAGCTGCTCACCGAGGCC 1293

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257
Db       1294 ATGGACTGGCCGGACAAGGGCGACGGCGGACTGCGCAGGGCCGCGGTCTCCTCCTTCGGC 1353

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
Db       1354 GTCAGCGGGACGAACGCGCACGTCGTGCTCGAAGAGGCCCCGGCG 1398
```

```
RESULT 13
US-09-105-537-34
; Sequence 34, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-34

  Query Match             48.0%;  Score 625.4;  DB 4;  Length 4689;
  Best Local Similarity   68.6%;  Pred. No. 2.8e-114;
  Matches  895;  Conservative    0;  Mismatches  401;  Indels    9;  Gaps    2;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
Db        103 GAGCCGGTGGCGATCGTGGGCATGGCCTGCCGCCTGCCGGGCGGTGTCGCCTCGCCCGAG 162

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
```

```
Db        163 GACCTGTGGCAGCTGGTGGCCGGGGACGGGGACGCGATCTCGGAGTTCCCGCAGGACCGC 222

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
Db        223 GGCTGGGACGTGGAGGGGCTGTACGACCCCGACCCGGACGCGTCCGGCAGGACGTACTGC 282

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
Db        283 CGGTCCGGCGGATTCCTGCACGACGCCGGCGAGTTCGACGCCGACTTCTTCGGGATCTCG 342

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
Db        343 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGACTGTCCCTCACCACCGCGTGGGAG 402

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
Db        403 GCGATCGAGAGCGCGGGCATCGACCCGACGGCCCTGAAGGGCAGCGGCCTCGGCGTCTTC 462

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
Db        463 GTCGGCGGCTGGCACACCGGCTACACCTCGGGGCAGACCACCG------CCGTGCAGTCG 516

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
Db        517 CCCGAGCTGGAGGGCCACCTGGTCAGCGGCGCGGCGCTGGGCTTCCTGTCCGGCCGTATC 576

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
Db        577 GCGTACGTCCTCGGTACGGACGGACCGGCCCTGACCGTGGACACGGCCTGCTCGTCCTCG 636

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
Db        637 CTGGTCGCCCTGCACCTCGCCGTGCAGGCCCTCCGCAAGGGCGAGTGCGACATGGCCCTC 696

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
Db        697 GCCGGTGGTGTCACGGTCATGCCCAACGCGGACCTGTTCGTGCAGTTCAGCCGGCAGCGC 756

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
Db        757 GGGCTGGCCGCGGACGGCCGGTCGAAGGCGTTCGCCACCTCGGCGGACGGCTTCGGCCCC 816

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
Db        817 GCGGAGGGCGCCGGAGTCCTGCTGGTGGAGCGCCTGTCGGACGCCCGCCGCAACGGACAC 876

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
Db        877 CGGATCCTCGCGGTCGTCCGCGGCAGCGCGGTCAACCAGGACGGCGCCAGCAACGGCCTC 936

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
Db        937 ACGGCTCCGCACGGGCCCTCCCAGCAGCGCGTCATCCGACGGGCCCTGGCGGACGCCCGG 996

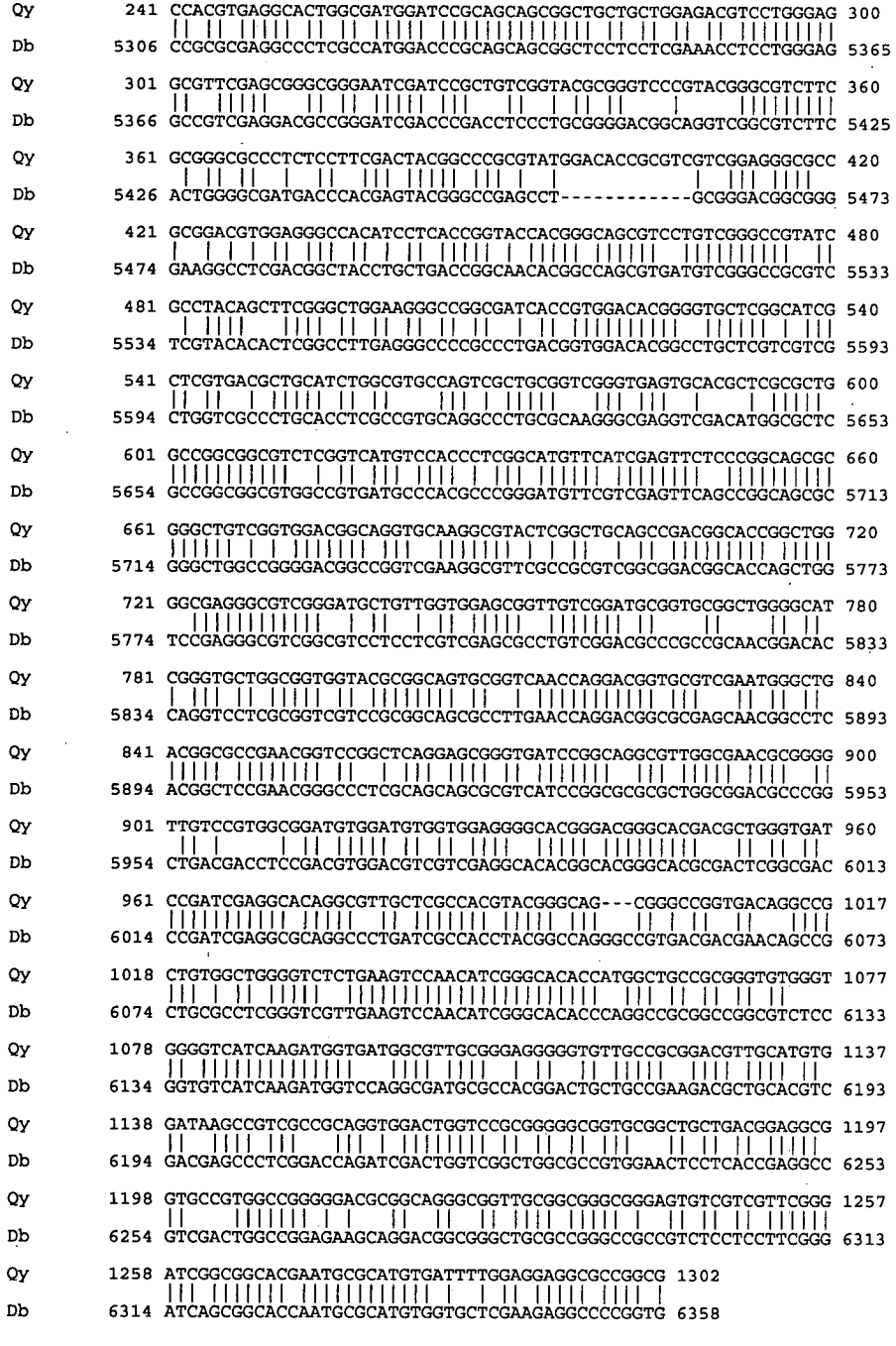
Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
Db        997 CTCGCGCCGGGTGACGTGGACGTCGTCGAGGCGCACGGCACGGGCACGCGGCTCGGCGAC 1056

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGG---GCAGCGGGCCGGTGACAGGCCG 1017
Db       1057 CCGATCGAGGCGCAGGCCCTCATCGCCACCTACGGCCAGGAGAAGAGCAGCGAACAGCCG 1116

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077
Db       1117 CTGAGGCTGGGCGCGTTGAAGTCGAACATCGGGCACACGCAGGCCGCGGCCGGTGTCGCA 1176

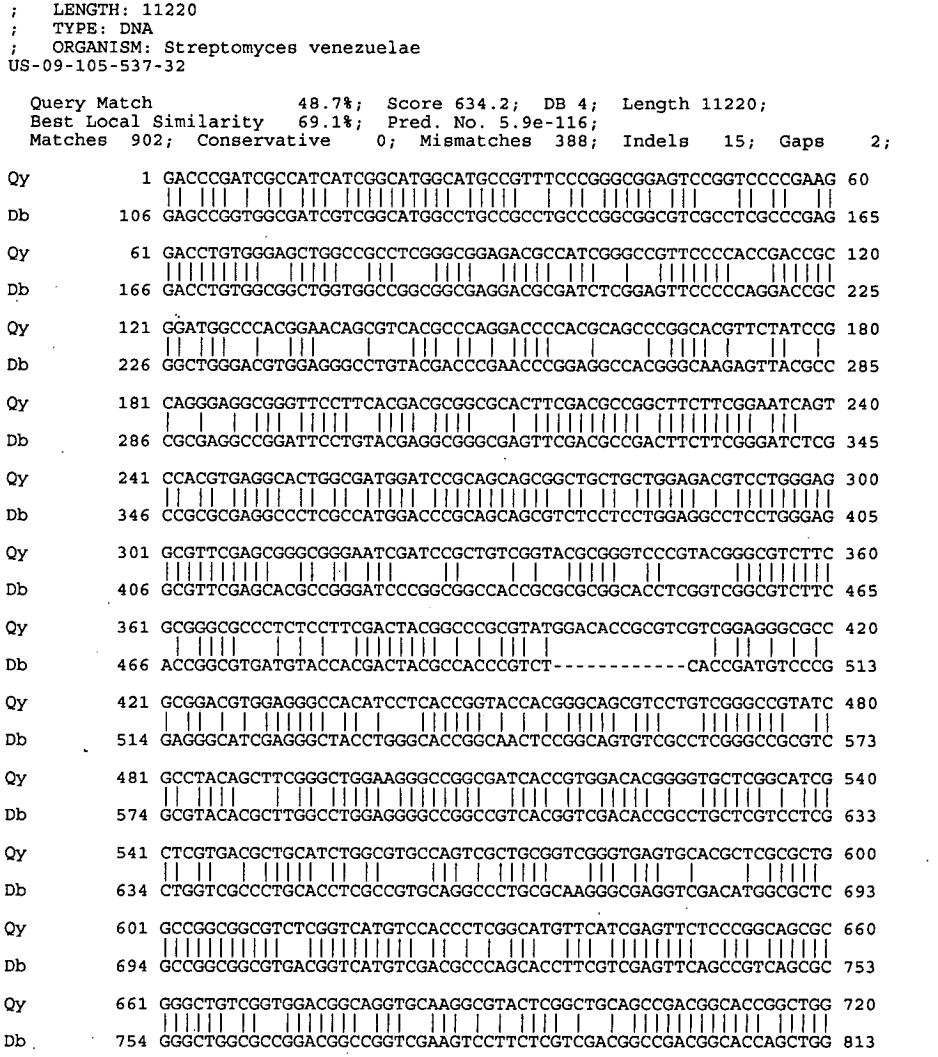
Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137
Db       1177 GGTGTCATCAAGATGGTCCAGGCGATGCGCCACGGACTGCTGCCGAAGACGCTGCACGTC 1236

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197
Db       1237 GACGAGCCCTCGGACCAGATCGACTGGTCGGCGGGCACGGTGGAACTCCTCACCGAGGCC 1296
```



```
RESULT 12
US-09-105-537-32
; Sequence 32, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-32

  Query Match             48.7%;  Score 634.2;  DB 4;  Length 11220;
  Best Local Similarity   69.1%;  Pred. No. 5.9e-116;
  Matches  902;  Conservative    0;  Mismatches  388;  Indels   15;  Gaps    2;
```



```
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-30

  Query Match             49.1%;  Score 639;  DB 4;  Length 13842;
  Best Local Similarity   69.3%;  Pred. No. 7e-117;
  Matches  905;  Conservative    0;  Mismatches  385;  Indels   15;  Gaps    2;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || ||| | || ||| ||||||||||| |||||  | || || || |||   || ||| ||
Db       3325 GAGCCGGTGGCGATCGTCGGCATGGCCTGCCGCCTGCCCGGTGGGGTCGCCTCGCCGGAG 3384

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              |||||||||  |||||  |||   ||||| ||||| |||  |  ||||||    ||||||
Db       3385 GACCTGTGGCGGCTGGTGGCCGGCGGCGGGGACGCGATCTCGGAGTTCCCGCAGGACCGC 3444

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||  |  |||  ||  ||  || ||| ||  ||| ||||||||||  ||
Db       3445 GGCTGGGACGTGGAGGGGCTGTACCACCCGGATCCGGAGCACCCCGGCACGTCGTACGTC 3504

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              |     ||||| ||| |  |  |||  ||   ||||||||| | |||||||| |||
Db       3505 CGCCAGGGCGGTTTCATCGAGAACGTCGCCGGCTTCGACGCGGCCTTCTTCGGGATCTCG 3564

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              || || ||||| || || ||||| ||||||||||||| || || || || ||||||||||
Db       3565 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTCGAAACCTCCTGGGAG 3624

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||  |||||   || || ||||| |||   ||  | || ||     |      |||||||||
Db       3625 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTGCGGGGACGGCAGGTCGGCGTCTTC 3684

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              | || ||  |  ||  ||| ||||| ||| |  |           |  ||| ||||
Db       3685 ACTGGGGCGATGACCCACGAGTACGGGCCGAGCCT-----------GCGGGACGGCGGG 3732

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              | | | | || ||| || | || ||||| | ||||| ||||||  ||||||||||  ||
Db       3733 GAAGGCCTCGACGGCTACCTGCTGACCGGCAACACGGCCAGCGTGATGTCGGGCCGCGTC 3792

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              | ||||   |||| || || || || ||  | || ||||||||||  |||||| | |||
Db       3793 TCGTACACACTCGGCCTTGAGGGCCCCGCCCTGACGGTGGACACGGCCTGCTCGTCGTCG 3852

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              || ||  | ||||| || ||     ||| | ||||||  ||| |||  |    | |||||
Db       3853 CTGGTCGCCCTGCACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAGGTCGACATGGCGCTC 3912

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              |||||||||||  | || ||| |||| | ||| |||||| ||||||| ||||||||||
Db       3913 GCCGGCGGCGTGGCCGTGATGCCCACGCCCGGGATGTTCGTCGAGTTCAGCCGGCAGCGC 3972

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              ||||||  |  |||||||  |||  |||||||  |  |  ||  || |||||||| |||||
Db       3973 GGGCTGGCCGGGGACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCACCAGCTGG 4032

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              |||||||||||||  |  ||  |  ||  |||||  |||||||  ||     ||  || ||
Db       4033 TCCGAGGGCGTCGGCGTCCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACAC 4092

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              | ||| || ||||| || |||||||| ||  | ||||||||||| |||   || || ||
Db       4093 CAGGTCCTCGCGGTCGTCCGCGGCAGCGCCTTGAACCAGGACGGCGCGAGCAACGGCCTC 4152

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||| |||||||| || | ||| |||| || |||||||   ||| ||||| ||||  ||
Db       4153 ACGGCTCCGAACGGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTGGCGGACGCCCGG 4212

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              || |      | || ||||| || || ||||  ||||| |||||||||   || || ||
Db       4213 CTGACGACCTCCGACGTGGACGTCGTCGAGGCACACGGCACGGGCACGCGACTCGGCGAC 4272

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAG---CGGGCCGGTGACAGGCCG 1017
              |||||||||| |||||  || |||||| ||||| |||   || | ||  ||  ||||
Db       4273 CCGATCGAGGCGCAGGCCCTGATCGCCACCTACGGCCAGGGCCGTGACGACGAACAGCCG 4332

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077
              ||| | || ||||| |||||||||||||||||||||||||  ||| || || || ||
Db       4333 CTGCGCCTCGGGTCGTTGAAGTCCAACATCGGGCACACCCAGGCCGCGGCCGGCGTCTCC 4392

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137
              || |||||||||||||    |||| |||  | ||  || |||||  |||| |||| ||
Db       4393 GGTGTCATCAAGATGGTCCAGGCGATGCGCCACGGACTGCTGCCGAAGACGCTGCACGTC 4452

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197
              ||  |||| |||   ||| | |||||||| || || || |||   || || || |||||
Db       4453 GACGAGCCCTCGGACCAGATCGACTGGTCGGCTGGCGCCGTGGAACTCCTCACCGAGGCC 4512

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257
              ||     |||||| | |   ||  ||  || |||| |||||  |  || || || ||||||
Db       4513 GTCGACTGGCCGGAGAAGCAGGACGGCGGGCTGCGCCGGGCCGCCGTCTCCTCCTTCGGG 4572

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              ||| |||||||. ||||||||||||  |  | || |||||  |||| |
Db       4573 ATCAGCGGCACCAATGCGCATGTGGTGCTCGAAGAGGCCCCGGTG 4617


RESULT 11
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-5

  Query Match             49.1%;  Score 639;  DB 4;  Length 36778;
  Best Local Similarity   69.3%;  Pred. No. 8e-117;
  Matches  905;  Conservative    0;  Mismatches  385;  Indels   15;  Gaps    2;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || ||| | || ||| |||||||||| |||||  | || || || |||   || ||| ||
Db       5066 GAGCCGGTGGCGATCGTCGGCATGGCCTGCCGCCTGCCCGGTGGGGTCGCCTCGCCGGAG 5125

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              |||||||||  |||||  |||   ||||| ||||| |||  |  ||||||    ||||||
Db       5126 GACCTGTGGCGGCTGGTGGCCGGCGGCGGGGACGCGATCTCGGAGTTCCCGCAGGACCGC 5185

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||  |  |||  ||  ||  || ||| ||  ||| ||||||||||  ||
Db       5186 GGCTGGGACGTGGAGGGGCTGTACCACCCGGATCCGGAGCACCCCGGCACGTCGTACGTC 5245

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              |     ||||| ||| |  |  |||  ||   ||||||||| | |||||||| |||
Db       5246 CGCCAGGGCGGTTTCATCGAGAACGTCGCCGGCTTCGACGCGGCCTTCTTCGGGATCTCG 5305
```

```
Db       4156 CCGATCGAGGCGCAGGCCCTGATCGCCACCTACGGCCAGGGCCGTGACGACGAACAGCCG 4215

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077

Db       4216 CTGCGCCTCGGGTCGTTGAAGTCCAACATCGGGCACACCCAGGCCGCGGCCGGCGTCTCC 4275

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137

Db       4276 GGTGTCATCAAGATGGTCCAGGCGATGCGCCACGGACTGCTGCCGAAGACGCTGCACGTC 4335

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197

Db       4336 GACGAGCCCTCGGACCAGATCGACTGGTCGGCTGGCGCCGTGGAACTCCTCACCGAGGCC 4395

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257

Db       4396 GTCGACTGGCCGGAGAAGCAGGACGGCGGGCTGCGCCGGGCCGCCGTCTCCTCCTTCGGG 4455

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302

Db       4456 ATCAGCGGCACCAATGCGCATGTGGTGCTCGAAGAGGCCCCGGTG 4500


RESULT 9
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-1

  Query Match             49.3%;  Score 641.6;  DB 4;  Length 15872;
  Best Local Similarity   70.1%;  Pred. No. 2.2e-117;
  Matches  914;  Conservative    0;  Mismatches  369;  Indels   21;  Gaps    3;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60

Db       8498 GATCCGATCGCGATCGTGGGCATGGCATGCCGCTATCCGGGTGGTGTGTCGTCGCCGGAG 8557

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120

Db       8558 GACCTGTGGCGGCTGGTGGCCGAGGGGACGGACGCGATCAGCGAGTTCCCCGTCAACCGC 8617

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180

Db       8618 GGCTGGGACCTGGAGAGCCTCTACGACCCGGATCCCGAGTCGAAGGGCACCACGTACTGC 8677

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240

Db       8678 CGGGAGGGCGGGTTCCTGGAAGGCGCCGGTGACTTCGACGCCGCCTTCTTCGGCATCTCG 8737

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300

Db       8738 CCGCGCGAGGCCCTGGTGATGGACCCGCAGCAGCGGCTGCTGCTGGAGGTGTCCTGGGAG 8797

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360

Db       8798 GCGCTGGAACGCGCGGGCATCGACCCGTCCTCGCTGCGCGGCAGCCGCGGTGGTGTCTAC 8857

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420

Db       8858 GTGGGCGCCGCGCACGGCTCGTAC----------GCCTCCGATCCCCGGCTGGTGCCC 8905

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480

Db       8906 GAGGGCTCGGAGGGCTATCTGCTGACCGGCAGCGCCGACGCGGTGATGTCCGGCCGCATC 8965

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540

Db       8966 TCCTACGCGCTCGGTCTCGAAGGACCGTCCATGACGGTGGAGACGGCCTGCTCCTCCTCG 9025

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600

Db       9026 CTGGTGGCGCTGCATCTGGCGGTACGGGCGCTGCGGCACGGCGAGTGCGGGCTCGCGCTG 9085

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660

Db       9086 GCGGGCGGGGTGGCGGTGATGGCCGATCCGGCGGCGTTCGTGGAGTTCTCCCGGCAGAAG 9145

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720

Db       9146 GGGCTGGCCGCCGACGGCCGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGGCTGG 9205

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780

Db       9206 GCCGAGGGCGTCGGCGTGCTCGTCCTGGAGCGGCTGTCGGACGCGCGCCGCGCGGGGCAC 9265

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840

Db       9266 ACGGTCCTCGGCCTGGTCACCGGCACCGCGGTCAACCAGGACGGTGCCTCCAACGGGCTG 9325

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900

Db       9326 ACCGCGCCCAACGGCCCAGCCCAGCAACGCGTCATCGCCGAGGCGCTCGCCGACGCCGGG 9385

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960

Db       9386 CTGTCCCCGGAGGACGTGGACGCGGTCGAGGCGCACGGCACCGGCACCCGGCTCGGCGAC 9445

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACG---GGCAGCGGGCCGGTGACAGGCCG 1017

Db       9446 CCCATCGAGGCCGGGGCGCTGCTCGCCGCCTCCGGACGGAACCGTTCCGGCGACCACCCG 9505

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077

Db       9506 CTGTGGCTCGGCTCGCTGAAGTCCAACATCGGGCATGCCCAGGCCGCCGCCGGTGTCGGC 9565

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137

Db       9566 GGCGTCATCAAGATGCTCCAGGCGCTGCGGCACGGCTTGCTGCCCCGCACCCTCCACGCC 9625

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197

Db       9626 GACGAGCCGACCCCGCATGCCGACTGGAGCTCCGGCCGGGTACGGCTGCTCACCTCCGAG 9685

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257

Db       9686 GTGCCGTGGCAGCG------GACCGGCCGGCCCCGGCGGACCGGGGTGTCCGCCTTCGGC 9739

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGC 1301

Db       9740 GTCGGCGGCACCAATGCCCATGTCGTCCTCGAAGAGGCACCCGC 9783


RESULT 10
US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
```

```
Db      50969 GGTCTGGCTCCGGATGGGCGGTGCAAGTCGTTCGCGGCGGCCGCGGATGGCACCGGGTGG 51028

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780

Db      51029 GGCGAGGGTGCCGGTCTGGTGTTGCTGGAGCGGTTGTCGGATGCCCGGCGCAATGGGCAC 51088

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840

Db      51089 GAGGTTCTGGCGGTGGTGCGGGGTAGCGCGGTGAACCAGGACGGCGCGTCGAATGGCTTG 51148

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900

Db      51149 ACTGCGCCAAATGGTCCGTCACAGCAAAGGGTGATCACCCAGGCACTCACGAGTGCCGGG 51208

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960

Db      51209 CTGTCCGTGTCCGACGTGGATGCTGTGGAGGCGCATGGGACGGGCACGCGGCTTGGTGAT 51268

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACG---GGCAGCGGGCCGGTGACAGGCCG 1017

Db      51269 CCGATCGAGGCGCAGGCGTTGATCGCTACGTACGGCCGGGATCGTGATCCCGGTCGGCCG 51328

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077

Db      51329 TTGTGGCTGGGGTCGGTGAAGTCGAATATTGGTCACACCCAGGCGGCGGCGGGTGTCGCT 51388

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137

Db      51389 GGTGTGATCAAGATGGTGATGGCGATGCGGCAGGGGGAGCTGCCGCGCACGTTGCACGTG 51448

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197

Db      51449 GACGAGCCCTCCGCGCAGGTGGACTGGTCTGCGGGCACGGTCCAACTCCTCACGGAGAAC 51508

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257

Db      51509 ACGCCCTGGCC------CGACAGCGGTCGTCTTCGCCGGGCGGGCGTGTCATCGTTCGGG 51562

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCG 1299

Db      51563 ATCAGTGGCACCAACGCGCACCTGATCCTTGAACAACCTCCG 51604


RESULT 8
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-19

  Query Match             49.3%;  Score 642.2;  DB 3;  Length 38506;
  Best Local Similarity   69.5%;  Pred. No. 1.9e-117;
  Matches  907;  Conservative    0;  Mismatches  383;  Indels   15;  Gaps    2;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60

Db       3208 GAGCCGGTGGCGATCGTCGGCATGGCCTGCCGCCTGCCCGGTGGGGTCGCCTCGCCGGAG 3267

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120

Db       3268 GACCTGTGGCGGCTGGTGGCCGGCGGCGGGGACGCGATCTCGGAGTTCCCGCAGGACCGC 3327

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180

Db       3328 GGCTGGGACGTGGAGGGGCTGTACCACCCGGATCCCGAGCACCCCGGCACGTCGTACGTC 3387

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240

Db       3388 CGCCAGGGCGGTTTCATCGAGAACGTCGCCGGCTTCGACGCGGCCTTCTTCGGGATCTCG 3447

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300

Db       3448 CCGCGCGAGGCCCTCGCCATGGACCCGCAGCAGCGGCTCCTCCTCGAAACCTCCTGGGAG 3507

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360

Db       3508 GCCGTCGAGGACGCCGGGATCGACCCGACCTCCCTGCGGGGACGGCAGGTCGGCGTCTTC 3567

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420

Db       3568 ACTGGGGCGATGACCCACGAGTACGGGCCGAGCCT----------GCGGGACGGCGGG 3615

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480

Db       3616 GAAGGCCTCGACGGCTACCTGCTGACCGGCAACACGGCCAGCGTGATGTCGGGCCGCGTC 3675

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540

Db       3676 TCGTACACACTCGGCCTTGAGGGCCCCGCCCTGACGGTGGACACGGCCTGCTCGTCGTCG 3735

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600

Db       3736 CTGGTCGCCCTGCACCTCGCCGTGCAGGCCCTGCGCAAGGGCGAGGTCGACATGGCGCTC 3795

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660

Db       3796 GCCGGCGGCGTGGCCGTGATGCCCACGCCCGGGATGTTCGTCGAGTTCAGCCGGCAGCGC 3855

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720

Db       3856 GGGCTGGCCGGGGACGGCCGGTCGAAGGCGTTCGCCGCGTCGGCGGACGGCACCAGCTGG 3915

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780

Db       3916 TCCGAGGGCGTCGGCGTCCTCCTCGTCGAGCGCCTGTCGGACGCCCGCCGCAACGGACAC 3975

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840

Db       3976 CAGGTCCTCGCGGTCGTCCGCGGCAGCGCCGTGAACCAGGACGGCGCGAGCAACGGCCTC 4035

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900

Db       4036 ACGGCTCCGAACGGGCCCTCGCAGCAGCGCGTCATCCGGCGCGCGCTGGCGGACGCCCGG 4095

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960

Db       4096 CTGACGACCTCCGACGTGGACGTCGTCGAGGCACACGGCACGGGCACGCGACTCGGCGAC 4155

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAG---CGGGCCGGTGACAGGCCG 1017
```

```
Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
Db      50618 ACCTTGGAGCGGGCTGGGATCGATCCGTTCTCGTTGCACGGCAGCCGGACCGGTGTGTTC 50677

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
Db      50678 GCGGGCTTGATGTACCACGACTATGGGGCCCGATTCATTACCAGAGCACCGGAGGGCTTC 50737

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
Db      50738 ---------GAAGGGCACCTCGGGACGGGCAATGCGGGGAGCGTGCTGTCGGGTCGGGTT 50788

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
Db      50789 GCGTATTCGTTTGGTTTCGAGGGTCCTGCGGTGACGGTGGATACGGCGTGTTCGTCGTCG 50848

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
Db      50849 TTGGTGGCGTTACACCTGGCGGGTCAAGCACTGCGGGCCGGTGAGTGCGAATTCGCCCTT 50908

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
Db      50909 GCCGGTGGCGTCACGGTGATGTCGACGCCGACGACGTTCGTGGAGTTCTCCCGTCAACGG 50968

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
Db      50969 GGTCTGGCTCCGGATGGGCGGTGCAAGTCGTTCGCGGCGGCCGCGGATGGCACCGGGTGG 51028

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
Db      51029 GGCGAGGGTGCCGGTCTGGTGTTGCTGGAGCGGTTGTCGGATGCCCGGCGCAATGGGCAC 51088

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
Db      51089 GAGGTTCTGGCGGTGGTGCGGGGTAGCGCGGTGAACCAGGACGGCGCGTCGAATGGCTTG 51148

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
Db      51149 ACTGCGCCAAATGGTCCGTCACAGCAAAGGGTGATCACCCAGGCACTCACGAGTGCCGGG 51208

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
Db      51209 CTGTCCGTGTCCGACGTGGATGCTGTGGAGGCGCATGGGACGGGCACGCGGCTTGGTGAT 51268

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACG---GGCAGCGGGCCGGTGACAGGCCG 1017
Db      51269 CCGATCGAGGCGCAGGCGTTGATCGCTACGTACGGCCGGGATCGTGATCCCGGTCGGCCG 51328

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077
Db      51329 TTGTGGCTGGGGTCGGTGAAGTCGAATATTGGTCACACCCAGGCGGCGGCGGGTGTCGCT 51388

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137
Db      51389 GGTGTGATCAAGATGGTGATGGCGATGCGGCAGGGGGAGCTGCCGCGCACGTTGCACGTG 51448

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197
Db      51449 GACGAGCCCTCCGCGCAGGTGGACTGGTCTGCGGGCACGGTCCAACTCCTCACGGAGAAC 51508

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257
Db      51509 ACGCCCTGGCC------CGACAGCGGTCGTCTTCGCCGGGCGGGCGTGTCATCGTTCGGG 51562

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCG 1299
Db      51563 ATCAGTGGCACCAACGCGCACCTGATCCTTGAACAACCTCCG 51604

RESULT 7
US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;    LENGTH: 80161
;    TYPE: DNA
;    ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

  Query Match           50.0%;  Score 650.4;  DB 4;  Length 80161;
  Best Local Similarity  70.5%;  Pred. No. 5.2e-119;
  Matches  918;  Conservative    0;  Mismatches  366;  Indels   18;  Gaps    3;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
Db      50318 GATCTTGTTGCGATTGTCGGGATGGGATGTCGTTTTCCGGGTGGGGTGTCGTCGCCGGAG 50377

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
Db      50378 GAGCTTTGGCGGCTGGTGGCCGGCGGCGTGGATGCGGTGGCTGGGTTCCCAGACGATCGC 50437

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
Db      50438 GGCTGGGATCTCGCGGCGTTGTACGATCCTGATCCCGATCGTCTCGGGACCTCGTATGTG 50497

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
Db      50498 TGTGAGGGCGGGTTTCTGCGGGACGCGGCGGAGTTCGATGCTGACATGTTCGGCATCAGC 50557

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
Db      50558 CCGCGTGAGGCGTTGGCGATGGATCCGCAGCAGCGGTTGCTGCTGGAGGTCGCCTGGGAA 50617

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
Db      50618 ACCTTGGAGCGGGCTGGGATCGATCCGTTCTCGTTGCACGGCAGCCGGACCGGTGTGTTC 50677

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
Db      50678 GCGGGCTTGATGTACCACGACTATGGGGCCCGATTCATTACCAGAGCACCGGAGGGCTTC 50737

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
Db      50738 ---------GAAGGGCACCTCGGGACGGGCAATGCGGGGAGCGTGCTGTCGGGTCGGGTT 50788

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
Db      50789 GCGTATTCGTTTGGTTTCGAGGGTCCTGCGGTGACGGTGGATACGGCGTGTTCGTCGTCG 50848

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
Db      50849 TTGGTGGCGTTACACCTGGCGGGTCAAGCACTGCGGGCCGGTGAGTGCGAATTCGCCCTT 50908

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
Db      50909 GCCGGTGGCGTCACGGTGATGTCGACGCCGACGACGTTCGTGGAGTTCTCCCGTCAACGG 50968

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
```

```
Db      31517 CGGGAGGGCGGCTTCCTCCACGACGCGGGCGACTTCGACGCGGACTTCTTCGGGATTTCG 31576

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300

Db      31577 CCGCGTGAGGCGGTGGCGATGGATCCGCAGCAGCGGCTGTTGCTGGAGACCTCCTGGGAG 31636

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360

Db      31637 GCGATCGAAGCGGCCGGTATCGACCCGCGAGGACTCCGCGGCAGCCGCACCGGGGTGTAC 31696

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420

Db      31697 GTGGGCGCCTGGGACAGCGGCTACACCGGCCAGGCGCACGCGCCCTCG---------- 31744

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480

Db      31745 GCCGAGTTGGAGGCCGACCTGCTGACCGGCGGCGTCGTCAGCTTCACCTCCGGCCGTATC 31804

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540

Db      31805 GCCTACACGCTGGGGCTGGAGGGTCCGGCCTTGACCGTGGACACCGCGTGTTCGTCGTCG 31864

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600

Db      31865 CTGGTCGCCCTGCACAACGCGGCGCAGGCGCTGCGGCGCGGCGAATGCGACCTGGCGTTG 31924

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660

Db      31925 GCCGGTGGTGTGACGGTGATGGCGACCCCGGCGGTGTTCGTCCAGTTCGCCCGGCAGCGA 31984

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720

Db      31985 GGGCTGGCGCCGGACGGCCGCTGCAAGGCGTTCGCCGACGCCGCCGACGGCTTCGGCCCC 32044

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780

Db      32045 GCCGAGGGTGTGGGGATGGTGCTGGTGGAGCGGTTGTCGGATGCCCGGCGGTTGGGGCAT 32104

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840

Db      32105 CCGGTGTTGGCGGTGGTGTGTGGGTCGGCGGTGAATCAGGACGGTGCGTCGAATGGTTTG 32164

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900

Db      32165 ACGGCGCCGAGTGGTCCGTCGCAGGAGCGGGTGATTCGTCAGGCGTTGGGGAATGCGCGG 32224

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960

Db      32225 TTGACGGTGGCGGATGTGGATGTGGTGGAGGCGCATGGGACGGGGACGCGGCTGGGTGAT 32284

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGC---AGCGGGCCGGTGACAGGCCG 1017

Db      32285 CCGATCGAGGCGCAGGCGTTGCTGGGGACGTATGGGCGGGATCGTGATGGTGGGCGTCCG 32344

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077

Db      32345 GTGTGGTTGGGGTCGTTGAAGTCGAATATTGGTCATGCTCAGGCGGCTGCGGGGGTGGCT 32404

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137

Db      32405 GGTGTGATCAAGATGGTGTTGGCGATGCGGTATGGGTGGTTGCCGCGGACGTTGCATGTG 32464

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197

Db      32465 GATGAGCCGAGCCGGCATGTGGACTGGTCGGCTGGTGGTGTGCGGTTGCTGACCGAGGCG 32524

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257

Db      32525 CGGGAGTGGCCGGG------GGTGGACCGGCCGCGTCGGGCGGCGGTCTCCGCCTTTGGT 32578

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGG 1291

Db      32579 GTCAGTGGTACCAACGCCCATCTGATCCTCGAAG 32612
```

```
RESULT 6
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  80161 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-036-987A-1

  Query Match           50.0%;  Score 650.4;  DB 3;  Length 80161;
  Best Local Similarity  70.5%;  Pred. No. 5.2e-119;
  Matches  918;  Conservative    0;  Mismatches  366;  Indels   18;  Gaps    3;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60

Db      50318 GATCTTGTTGCGATTGTCGGGATGGGATGTCGTTTTCCGGGTGGGGTGTCGTCGCCGGAG 50377

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120

Db      50378 GAGCTTTGGCGGCTGGTGGCCGGCGGCGTGGATGCGGTGGCTGGGTTCCCAGACGATCGC 50437

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180

Db      50438 GGCTGGGATCTCGCGGCGTTGTACGATCCTGATCCCGATCGTCTCGGGACCTCGTATGTG 50497

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240

Db      50498 TGTGAGGGCGGGTTTCTGCGGGACGCGGCGGAGTTCGATGCTGACATGTTCGGCATCAGC 50557

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300

Db      50558 CCGCGTGAGGCGTTGGCGATGGATCCGCAGCAGCGGTTGCTGCTGGAGGTCGCCTGGGAA 50617
```

```
Db      13647 CCGCGTGAGGCGTTGGCGATGGACCCGCAGCAGCGGTTGCTGCTGGAGACGTCCTGGGAG 13706

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360

Db      13707 GCCCTCGAACGGGCCGGCATCACTCCTTCCAAGCTGGCGGGCAGTCCGACCGGTGTGTTC 13766

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420

Db      13767 TTCGGCATGTCGAACCAGGACTACGCCGCCCAGGCGGGCGACGTGCCGTCCGAG------ 13820

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480

Db      13821 ------CTGGAGGGCTACCTGCTCACCGGCTCCATCTCCAGCGTCGCTTCGGGGCGTGTT 13874

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540

Db      13875 GCTTACACGTTCGGTCTTGAGGGGCCTGCGGTGACGGTGGATACGGCGTGTTCGTCGTCG 13934

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600

Db      13935 TTGGTGGCGTTGCATCTGGCGGTGCAGGGGTTGCGGCGGGGTGAGTGTTCGCTTGCGTTG 13994

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660

Db      13995 GTGGGTGGGGTGACGGTGATGTCGTCGCCGGTGACGTTGACGACGTTCAGTCGGCAGCGG 14054

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720

Db      14055 GGTTTGTCGGTGGATGGGCGGTGCAAGGCGTTCGCGGCTTCGGCGGATGGTTTTGGTGCT 14114

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780

Db      14115 GCCGAGGGTGTGGGTGTGTTGTTGGTGGAGCGGTTGTCGGATGCGCGGCGGTTGGGTCAT 14174

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840

Db      14175 CGGGTGTTGGCGGTGGTGCGGGGGAGTGCGGTCAATCAGGATGGTGCGTCCAATGGTCTG 14234

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900

Db      14235 GCGGCGCCGAATGGTCCGTCGCAGCAGCGGGTGATCCGTGCGGCGTTGGCTGACGCGGGT 14294

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960

Db      14295 CTGGCTCCTGCCGATGTGGATGTGGTGGAGGCGCATGGCACGGGGACGCGGTTGGGTGAT 14354

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAG---CGGGCCGGTGACAGGCCG 1017

Db      14355 CCGATCGAGGCTCAGGCGTTGCTGGCGACGTATGGGCAGGGTCGTACCAGTGGGCGTCCG 14414

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077

Db      14415 GTGTGGCTGGGGTCGGTGAAGTCGAACATCGGGCATACGCAGGCGGCGGCCGGTGTGGCT 14474

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137

Db      14475 GGTGTGATGAAGATGGTGCTGGCGTTGGGTCGGGGTGTGGTGCCGAAGACGTTGCATGTG 14534

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197

Db      14535 GATGAGCCGTCACCGCATGTGGACTGGTCGGCTGGTGAGGTGGAGTTGGCGGTTGAGGCG 14594

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257

Db      14595 GTGCCGTGGTCGCG------GGGTGGGCGGGTGCGGCGGGCTGGTGTGTCGTCGTTCGGG 14648

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302

Db      14649 ATCAGCGGCACGAATGCGCATGTGATCGTGGAGGAGGCGCCTGCG 14693

RESULT 5
US-08-804-227C-1

; Sequence 1, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  43280 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  816..14234
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14351..19945
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20010..31199
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31232..36067
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36249..41774
US-08-804-227C-1

  Query Match           52.0%;  Score 676.4;  DB 2;  Length 43280;
  Best Local Similarity  72.0%;  Pred. No. 3.9e-124;
  Matches  932;  Conservative    0;  Mismatches  341;  Indels   21;  Gaps    3;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60

Db      31337 GAGCCGATCGCCGTCGTCGGCATGGCCTGCCGCTACCCCGGAGGCGCGAACACCCCCGAA 31396

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120

Db      31397 CAGTTCTGGGAACTGCTCGACACCGGCACCGACGCCGCCGCGCCGATGCCCTCCGACCGG 31456

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180

Db      31457 GGATGGGACACCCACGGGCTGTACGACCCCGACCCGGCGGCAGCGGGGCGCACCTACTGC 31516

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
```

```
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-198-1

  Query Match            54.1%;  Score 704.8;  DB 2;  Length 44377;
  Best Local Similarity  72.6%;  Pred. No. 1.1e-129;
  Matches  945;  Conservative    0;  Mismatches  342;  Indels   15;  Gaps    2;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || ||||||| ||| | |||||||| ||||||| ||||||||| ||    ||| ||| |
Db       8270 GAGCCGATCGCGATCGTGGGCATGGCGTGCCGTTACCCGGGCGGTGTGGCGTCGCCGGAC 8329

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              |||||||||| |||| |||  | ||| || | ||  ||   |||||||  |||||||
Db       8330 GACCTGTGGGACCTGGTGGCCGGTGACGGGCACACGCTCTCCCCGTTCCCGGCCGACCGT 8389

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||  |  ||  ||   ||| || ||| ||   | ||| |||| |  ||||
Db       8390 GGCTGGGACGTCGAGGGGCTGTACGACCCGGAGCCGGGGGTGCCGGGCAAGAGCTATGTA 8449

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              | || |||||||||||  |   ||||||  | ||||||| |  |||||||| ||
Db       8450 CGGGAAGGCGGGTTCCTGCGTTCCGCGGCCGAGTTCGACGCGGAGTTCTTCGGGATATCG 8509

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              || || |||||   ||| ||||| |||||||||||| |||||||||||||| ||||||
Db       8510 CCGCGCGAGGCCACGGCCATGGACCCGCAGCAGCGGTTGCTGCTGGAGACGTCGTGGGAG 8569

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||| | |||||||| || |||| ||||   ||| | |||||  |||| || || ||||||
Db       8570 GCGCTGGAGCGGGCCGGCATCGTTCCGGACTCGCTGCGCGGCACCCGGACCGGTGTCTTC 8629

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              |||  |  |      ||||||           |  |||  |  |  ||| | ||||
Db       8630 AGCGGCATCTCCCAGCAGGACTAC------------GCGACCCAGCTGGGGGACGCCGCC 8677

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              |    |   |  || ||  | ||||| || |||   ||||| ||  | || || ||  |
Db       8678 GACACCTACGGCGGGCATGTGCTCACGGGGACCCTCGGCAGTGTGATCTCCGGTCGGGTT 8737

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              |||||    || ||| |||| |||||||| | || |||||||||| ||| ||| | |||
Db       8738 GCCTATGCGTTGGGGTTGGAGGGGCCGGCGCTGACGGTGGACACGGCGTGTTCGTCGTCG 8797

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              | ||| || |||||||||||   |||||| |||||  |||||||||    || ||| ||
Db       8798 TTGGTGGCGTTGCATCTGGCGGTGCAGTCGTTGCGGCGGGGTGAGTGTGATCTGGCGTTG 8857

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              ||||| || ||  |||| ||| | || |      ||||| | |||| ||||||||
Db       8858 GCCGGTGGGGTGACGGTGATGGCGACGCCGACGGTGTTCGTGGAGTTCTCGCGGCAGCGG 8917

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              |||||| ||| ||||||  |||||||||||| | |||  |  || |||||| || |  |||
Db       8918 GGGCTGGCGGCGGACGGGCGGTGCAAGGCGTTCGCGGAGGGTGCGGACGGGACGGCGTGG 8977

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              |  ||||| || ||  ||||| |||||||||| | || || |||   ||   || |||
Db       8978 GCGGAGGGTGTGGGTGTGCTGCTGGTGGAGCGGCTTTCCGACGCGCGCCGCAACGGTCAT 9037

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              |||||||||||||||| || |||||||||||||||| ||||||||||  |||||||||
Db       9038 CGGGTGCTGGCGGTGGTGCGGGGCAGTGCGGTCAATCAGGACGGTGCGAGCAATGGGCTG 9097

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||||||||  |||||||| ||| ||||||||||||||   ||||| ||||  | ||||||
Db       9098 ACGGCGCCGAGTGGTCCGGCGCAGCAGCGGGTGATCCGTGAGGCGCTGGCTGATGCGGGG 9157

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||      || || ||||||||||||||| |||||| ||||| ||| ||||||||||
Db       9158 CTGGTGCCCGCCGACGTGGATGTGGTGGAGGCGCACGGTACGGGGACGGCGCTGGGTGAT 9217

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              |||||||||||    ||| |||| |||||||||||||  |   |   | | |   ||| ||
Db       9218 CCGATCGAGGCGGGTGCGCTGCTGGCCACGTACGGGCGGGAGCGGGTCGGCGATCCGTTG 9277

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              ||||| |||||  ||||||| ||||||||||  |   ||| || ||||||||||||||
Db       9278 TGGCTCGGGTCGTTGAAGTCGAACATCGGGCATGCGCAGGCGGCTGCGGGTGTGGGTGGT 9337

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              || |||||| |||||  || | ||||| | |||  |||||||||||||| ||||||||||
Db       9338 GTGATCAAGGTGGTGCAGGGGATGCGGCATGGGTCGTTGCCGCGGACGCTGCATGTGGAT 9397

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              |||||||| || ||||||| ||| |  |||| ||||||  |||||||| ||| |   |
Db       9398 GCGCCGTCGTCGAAGGTGGAGTGGGCTTCGGGTGCGGTGGAGCTGCTGACCGAGACCCGG 9457

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              |||||||| ||     ||  | |||| |||||||||  |  |||||| ||||||| |
Db       9458 TCGTGGCCGCGG---CGGGTGGAGCGGGTGCGGCGGGCCGCGGTGTCGGCGTTCGGGGTG 9514

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              |||| || || || |||||| |  ||||||| |||||||||
Db       9515 AGCGGGACCAACGCCCATGTGGTCCTGGAGGAAGCGCCGGCG 9556


RESULT 4
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Shah, Sanjay Krishnakant
;  APPLICANT: McDaniel, Robert
;  APPLICANT: Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/428,517
;  CURRENT FILING DATE: 1999-10-28
;  EARLIER APPLICATION NUMBER: 60/120,254
;  EARLIER FILING DATE: 1999-02-16
;  EARLIER APPLICATION NUMBER: 60/106,100
;  EARLIER FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 50937
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artifical Sequence: Recombinant DNA
US-09-428-517-1

  Query Match            52.9%;  Score 689;  DB 4;  Length 50937;
  Best Local Similarity  72.3%;  Pred. No. 1.4e-126;
  Matches  944;  Conservative    0;  Mismatches  340;  Indels   21;  Gaps    3;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || ||||||| ||| | || |||||||||| |||||||| ||   ||    |||||| ||
Db      13407 GAGCCGATCGCGATCGTGGGAATGGCATGCCGCTTCCCGGGAGGCATCGCTTCCCCGGAG 13466

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              || ||||||||  ||  | || | ||  |||| ||     | ||      |||||
Db      13467 GATCTGTGGGACGTGCTCACCGCTGGTGAGGACGTTGTCTCCGGACTGCCGCAGAACCGC 13526

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||  |  ||  ||   ||| || ||| ||   | ||| |||| |  ||||
Db      13527 GGGTGGGACTTGGGGCGCCTGTACGATCCCGATCCGGACCGGGCCGGTACGTCATACATG 13586

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              | | || | ||||| ||||| |||| | | |||||||| | |||||||| |||
Db      13587 CGTGAGGGTGCTTTCCTGCACGAGGCGGGGGAGTTCGACGCGGCCTTCTTCGGTATCTCG 13646

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              || ||||||||  |||||||||| |||||||||||| |||||||||||||||||||||||
```

```
  Best Local Similarity   72.6%;  Pred. No. 1.1e-129;
  Matches  945;  Conservative    0;  Mismatches  342;  Indels   15;  Gaps    2;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || |||||||| ||| | |||||||| ||||||| ||||||||| ||   ||| ||| |
Db       8270 GAGCCGATCGCGATCGTGGGCATGGCGTGCCGTTACCCGGGCGGTGTGGCGTCGCCGGAC 8329

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              |||||||||||| ||||  |||   | |||  || |  ||   |||||||  |||||||
Db       8330 GACCTGTGGGACCTGGTGGCCGGTGACGGGCACACGCTCTCCCCGTTCCCGGCCGACCGT 8389

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||  |   ||   ||  ||| || ||| ||   |  ||| |||| |  ||||
Db       8390 GGCTGGGACGTCGAGGGGCTGTACGACCCGGAGCCGGGGGTGCCGGGCAAGAGCTATGTA 8449

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              | || ||||||||||||  |    ||||||  | ||||||||  |  ||||||||  ||
Db       8450 CGGGAAGGCGGGTTCCTGCGTTCCGCGGCCGAGTTCGACGCGGAGTTCTTCGGGATATCG 8509

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              || ||  |||||   ||| |||||  |||||||||||| |||||||||||||||| ||||||
Db       8510 CCGCGCGAGGCCACGGCCATGGACCCGCAGCAGCGGTTGCTGCTGGAGACGTCGTGGGAG 8569

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||| | ||||||||  || |||| ||||   ||| | |||||  |||| || || ||||||
Db       8570 GCGCTGGAGCGGGCCGGCATCGTTCCGGACTCGCTGCGCGGCACCCGGACCGGTGTCTTC 8629

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
                 |||  | |      ||||||           | |||  | |  ||| | ||||
Db       8630 AGCGGCATCTCCCAGCAGGACTAC-----------GCGACCCAGCTGGGGGACGCCGCC 8677

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              |    |  |  || ||  | ||||| || |||    ||||| ||  | || || ||  |
Db       8678 GACACCTACGGCGGGCATGTGCTCACGGGGACCCTCGGCAGTGTGATCTCCGGTCGGGTT 8737

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              |||||    || ||| |||| |||||||| | || |||||||||| ||| ||| | |||
Db       8738 GCCTATGCGTTGGGGTTGGAGGGGCCGGCGCTGACGGTGGACACGGCGTGTTCGTCGTCG 8797

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              | ||| || |||||||||||   |||||| |||||  |||||||||    || ||| ||
Db       8798 TTGGTGGCGTTGCATCTGGCGGTGCAGTCGTTGCGGCGGGGTGAGTGTGATCTGGCGTTG 8857

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              ||||| || ||  |||| ||| | || |      |||||| | ||||||| ||||||||
Db       8858 GCCGGTGGGGTGACGGTGATGGCGACGCCGACGGTGTTCGTGGAGTTCTCGCGGCAGCGG 8917

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              |||||| ||| ||||||  |||||||||||| | |||  |  || ||||| || |. ||| -
Db       8918 GGGCTGGCGGCGGACGGGCGGTGCAAGGCGTTCGCGGAGGGTGCGGACGGGACGGCGTGG 8977

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              | ||||| || ||  |||| |||||||||||| | || || |||   ||    || |||
Db       8978 GCGGAGGGTGTGGGTGTGCTGCTGGTGGAGCGGCTTTCCGACGCGCGCCGCAACGGTCAT 9037

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              |||||||||||||||| || |||||||||||||| ||||||||||    |||||||||
Db       9038 CGGGTGCTGGCGGTGGTGCGGGGCAGTGCGGTCAATCAGGACGGTGCGAGCAATGGGCTG 9097

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||||||||  |||||||| ||| ||||||||||||  |||| ||||  | ||||||
Db       9098 ACGGCGCCGAGTGGTCCGGCGCAGCAGCGGGTGATCCGTGAGGCGCTGGCTGATGCGGGG 9157

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||       || || |||||||||||||||| |||||| ||||| ||| ||||||||||
Db       9158 CTGGTGCCCGCCGACGTGGATGTGGTGGAGGCGCACGGTACGGGGACGGCGCTGGGTGAT 9217

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              ||||||||||    ||| |||| |||||||||||||| | |   |  | |   ||| ||
Db       9218 CCGATCGAGGCGGGTGCGCTGCTGGCCACGTACGGGCGGGAGCGGGTCGGCGATCCGTTG 9277

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              ||||| |||||  ||||||| ||||||||||  |   ||| || ||||||||||||||

Db       9278 TGGCTCGGGTCGTTGAAGTCGAACATCGGGCATGCGCAGGCGGCTGCGGGTGTGGGTGGT 9337

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              || |||||| |||||  || | ||||| | |||  ||||||||||||| ||||||||||
Db       9338 GTGATCAAGGTGGTGCAGGGGATGCGGCATGGGTCGTTGCCGCGGACGCTGCATGTGGAT 9397

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              |||||||| || ||||||| ||| |   |||| ||||||  |||||||| ||| |   |
Db       9398 GCGCCGTCGTCGAAGGTGGAGTGGGCTTCGGGTGCGGTGGAGCTGCTGACCGAGACCCGG 9457

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              |||||||| ||      ||  | |||| |||||||||| |  |||||| ||||||| |
Db       9458 TCGTGGCCGCGG---CGGGTGGAGCGGGTGCGGCGGGCCGCGGTGTCGGCGTTCGGGGTG 9514

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              |||| || || || |||||| |  ||||||| |||||||||
Db       9515 AGCGGGACCAACGCCCATGTGGTCCTGGAGGAAGCGCCGGCG 9556


RESULT 3
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
```

```
Db       8255 GAGCCGATCGCGATCGTGGGCATGGCGTGCCGTTACCCGGGCGGTGTGGCGTCGCCGGAC 8314

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120

Db       8315 GACCTGTGGGACCTGGTGGCCGGTGACGGGCACACGCTCTCCCCGTTCCCGGCCGACCGT 8374

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180

Db       8375 GGCTGGGACGTCGAGGGGCTGTACGACCCGGAGCCGGGGGTGCCGGGCAAGAGCTATGTA 8434

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240

Db       8435 CGGGAAGGCGGGTTCCTGCGTTCCGCGGCCGAGTTCGACGCGGAGTTCTTCGGGATATCG 8494

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300

Db       8495 CCGCGCGAGGCCACGGCCATGGACCCGCAGCAGCGGTTGCTGCTGGAGACGTCGTGGGAG 8554

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360

Db       8555 GCGCTGGAGCGGGCCGGCATCGTTCCGGACTCGCTGCGCGGCACCCGGACCGGTGTCTTC 8614

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420

Db       8615 AGCGGCATCTCCCAGCAGGACTAC----------GCGACCCAGCTGGGGGACGCCGCC 8662

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480

Db       8663 GACACCTACGGCGGGCATGTGCTCACGGGGACCCTCGGCAGTGTGATCTCCGGTCGGGTT 8722

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540

Db       8723 GCCTATGCGTTGGGGTTGGAGGGGCCGGCGCTGACGGTGGACACGGCGTGTTCGTCGTCG 8782

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600

Db       8783 TTGGTGGCGTTGCATCTGGCGGTGCAGTCGTTGCGGCGGGGTGAGTGTGATCTGGCGTTG 8842

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660

Db       8843 GCCGGTGGGGTGACGGTGATGGCGACGCCGACGGTGTTCGTGGAGTTCTCGCGGCAGCGG 8902

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720

Db       8903 GGGCTGGCGGCGGACGGGCGGTGCAAGGCGTTCGCGGAGGGTGCGGACGGGACGGCGTGG 8962

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780

Db       8963 GCGGAGGGTGTGGGTGTGCTGCTGGTGGAGCGGCTTTCCGACGCGCGCCGCAACGGTCAT 9022

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840

Db       9023 CGGGTGCTGGCGGTGGTGCGGGGCAGTGCGGTCAATCAGGACGGTGCGAGCAATGGGCTG 9082

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900

Db       9083 ACGGCGCCGAGTGGTCCGGCGCAGCAGCGGGTGATCCGTGAGGCGCTGGCTGATGCGGGG 9142

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960

Db       9143 CTGGTGCCCGCCGACGTGGATGTGGTGGAGGCGCACGGTACGGGGACGGCGCTGGGTGAT 9202

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020

Db       9203 CCGATCGAGGCGGGTGCGCTGCTGGCCACGTACGGGCGGGAGCGGGTCGGCGATCCGTTG 9262

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080

Db       9263 TGGCTCGGGTCGTTGAAGTCGAACATCGGGCATGCGCAGGCGGCTGCGGGTGTGGGTGGT 9322

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140

Db       9323 GTGATCAAGGTGGTGCAGGGGATGCGGCATGGGTCGTTGCCGCGGACGCTGCATGTGGAT 9382

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200

Db       9383 GCGCCGTCGTCGAAGGTGGAGTGGGCTTCGGGTGCGGTGGAGCTGCTGACCGAGACCCGG 9442

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260

Db       9443 TCGTGGCCGCGG---CGGGTGGAGCGGGTGCGGCGGGCCGCGGTGTCGGCGTTCGGGGTG 9499

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302

Db       9500 AGCGGGACCAACGCCCATGTGGTCCTGGAGGAAGCGCCGGCG 9541


RESULT 2
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-227C-7

  Query Match             54.1%;  Score 704.8;  DB 2;  Length 44377;
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 56.4013 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_1441_2742
Perfect score: 1302
Sequence: 1 gacccgatcgccatcatcgg.....ttttggaggaggcgccggcg 1302

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	704.8	54.1	13987	2	US-08-804-227C-13	Sequence 13, Appl
	2	704.8	54.1	44377	2	US-08-804-227C-7	Sequence 7, Appli
	3	704.8	54.1	44377	2	US-08-804-198-1	Sequence 1, Appli
	4	689	52.9	50937	4	US-09-428-517-1	Sequence 1, Appli
	5	676.4	52.0	43280	2	US-08-804-227C-1	Sequence 1, Appli
	6	650.4	50.0	80161	3	US-09-036-987A-1	Sequence 1, Appli
	7	650.4	50.0	80161	4	US-09-370-700-1	Sequence 1, Appli
	8	642.2	49.3	38506	3	US-09-320-878-19	Sequence 19, Appl
	9	641.6	49.3	15872	4	US-09-105-537-1	Sequence 1, Appli
	10	639	49.1	13842	4	US-09-105-537-30	Sequence 30, Appl
	11	639	49.1	36778	4	US-09-105-537-5	Sequence 5, Appli
	12	634.2	48.7	11220	4	US-09-105-537-32	Sequence 32, Appl
	13	625.4	48.0	4689	4	US-09-105-537-34	Sequence 34, Appl
	14	621.4	47.7	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	15	621.4	47.7	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	16	608.2	46.7	20235	1	US-07-642-734C-3	Sequence 3, Appli
	17	608.2	46.7	20235	3	US-08-439-009A-3	Sequence 3, Appli
c	18	604	46.4	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	19	599.8	46.1	11219	1	US-07-642-734C-1	Sequence 1, Appli
	20	599.8	46.1	11219	3	US-08-439-009A-1	Sequence 1, Appli
c	21	596.4	45.8	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	22	594.2	45.6	4041	4	US-09-105-537-36	Sequence 36, Appl
	23	533.2	41.0	33529	4	US-09-144-085-3	Sequence 3, Appli
	24	500.6	38.4	68750	3	US-09-335-409-1	Sequence 1, Appli
	25	500.6	38.4	68750	4	US-09-568-102-1	Sequence 1, Appli
	26	500.6	38.4	68750	4	US-09-567-969-1	Sequence 1, Appli
	27	500.6	38.4	68750	4	US-09-568-480-1	Sequence 1, Appli
	28	500.6	38.4	68750	4	US-09-568-486-1	Sequence 1, Appli
	29	500.6	38.4	68750	4	US-09-568-472-1	Sequence 1, Appli
	30	500.6	38.4	68750	4	US-09-567-899-1	Sequence 1, Appli
	31	495.8	38.1	71989	4	US-09-443-501A-2	Sequence 2, Appli
	32	488	37.5	28958	1	US-08-258-261B-6	Sequence 6, Appli
	33	488	37.5	28958	1	US-08-456-837-6	Sequence 6, Appli
	34	488	37.5	28958	1	US-08-457-342-6	Sequence 6, Appli
	35	488	37.5	28958	1	US-08-457-646A-6	Sequence 6, Appli
	36	488	37.5	28958	1	US-08-458-076A-6	Sequence 6, Appli
	37	488	37.5	28958	1	US-08-764-233A-4	Sequence 4, Appli
	38	488	37.5	28958	1	US-08-457-335A-6	Sequence 6, Appli
	39	488	37.5	28958	1	US-08-729-214-6	Sequence 6, Appli
	40	488	37.5	28958	3	US-09-028-934-6	Sequence 6, Appli
	41	488	37.5	49377	1	US-08-764-233A-1	Sequence 1, Appli
c	42	460.4	35.4	1419	4	US-09-434-288-2	Sequence 2, Appli
	43	411.2	31.6	1462	4	US-09-434-288-4	Sequence 4, Appli
c	44	338.4	26.0	1434	4	US-09-434-288-3	Sequence 3, Appli
	45	321.6	24.7	751	3	US-09-010-809-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-804-227C-13
; Sequence 13, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rosteck, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCI(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13987 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..13987
US-08-804-227C-13

Query Match 54.1%; Score 704.8; DB 2; Length 13987;
Best Local Similarity 72.6%; Pred. No. 9.3e-130;
Matches 945; Conservative 0; Mismatches 342; Indels 15; Gaps 2;

Qy 1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 1669.03 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_1441_2742
Perfect score: 1302
Sequence: 1 gacccgatcgccatcatcgg.....:...ttttggaggaggcgccggcg 1302

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match	Length	DB	ID	Description
c	1	111	8.5	1161	14	BM912505	BM912505 AGENCOURT
	2	110.6	8.5	1138	17	AG152606	AG152606 Pan trogl
	3	110.2	8.5	1201	14	BQ063878	BQ063878 AGENCOURT
	4	109.4	8.4	1243	14	BM913931	BM913931 AGENCOURT
	5	108	8.3	931	13	BI416575	BI416575 hasp001xg
	6	108	8.3	1037	17	AG146864	AG146864 Pan trogl
	7	107.8	8.3	1089	14	BQ930965	BQ930965 AGENCOURT
	8	107.8	8.3	1288	14	BM809948	BM809948 AGENCOURT
	9	107.2	8.2	888	17	AZ189327	AZ189327 SP_1014_A
	10	107.2	8.2	918	17	AQ895329	AQ895329 HS_4832_A
	11	107	8.2	662	9	AI514012	AI514012 GH27052.5
	12	106.8	8.2	863	13	BI416699	BI416699 hasp001xp
	13	105.8	8.1	1015	13	BI416650	BI416650 hasp001xm
	14	105.8	8.1	1089	14	BQ939440	BQ939440 AGENCOURT
	15	105	8.1	1197	13	BI416470	BI416470 hasp001xa
c	16	104.8	8.0	1289	14	BQ884996	BQ884996 AGENCOURT
	17	104.6	8.0	927	13	BI416685	BI416685 hasp001xo
	18	104.2	8.0	939	14	BQ890606	BQ890606 AGENCOURT
	19	104.2	8.0	1222	17	BH146899	BH146899 ENTQM47TF
	20	104	8.0	1137	17	AG078502	AG078502 Pan trogl
c	21	104	8.0	1236	14	BM914619	BM914619 AGENCOURT
	22	103.8	8.0	855	9	AL572700	AL572700 AL572700
c	23	103.8	8.0	1006	10	BE034822	BE034822 ML04H02 M
c	24	103.8	8.0	1021	14	BQ441712	BQ441712 AGENCOURT
c	25	103.8	8.0	1089	13	BM468944	BM468944 AGENCOURT
c	26	103.4	7.9	952	12	BG441341	BG441341 GA__Ea001
	27	103.2	7.9	1127	14	BQ718629	BQ718629 AGENCOURT
c	28	103	7.9	960	14	BQ429353	BQ429353 AGENCOURT
	29	103	7.9	1005	17	BH148610	BH148610 ENTPZ67TR
	30	102.8	7.9	670	17	AG126228	AG126228 Pan trogl
c	31	102.4	7.9	625	13	BM091001	BM091001 ig21a10.y
	32	102.4	7.9	685	12	BG786336	BG786336 SEAUMC006
	33	102.4	7.9	798	13	BI416695	BI416695 hasp001xo
	34	101.8	7.8	897	13	BI416517	BI416517 hasp001xd
	35	101.8	7.8	1478	14	BM911600	BM911600 AGENCOURT
c	36	101.6	7.8	877	17	AQ788449	AQ788449 HS_3130_A
	37	101.4	7.8	1184	14	BM911719	BM911719 AGENCOURT
	38	101.2	7.8	1301	14	BQ067141	BQ067141 AGENCOURT
	39	101	7.8	1090	17	AG128386	AG128386 Pan trogl
c	40	101	7.8	1303	14	BM908878	BM908878 AGENCOURT
	41	100.8	7.7	715	9	AL570062	AL570062 AL570062
	42	100.8	7.7	817	17	AZ196702	AZ196702 SP_1032_B
c	43	100.8	7.7	897	17	AZ186654	AZ186654 SP_1007_B
	44	100.8	7.7	924	10	BE455162	BE455162 HVSMEh009
	45	100.6	7.7	884	17	AQ782679	AQ782679 HS_3183_B

ALIGNMENTS

RESULT 1
BM912505/c
LOCUS BM912505 1161 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6612272 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5474051 5', mRNA sequence.
ACCESSION BM912505
VERSION BM912505.1 GI:19362884
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1161)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1986 row: k column: 12
High quality sequence stop: 218.
FEATURES Location/Qualifiers
source 1. .1161

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5474051"
/clone_lib="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 53 a 844 c 107 g 88 t 69 others
ORIGIN

Query Match 8.5%; Score 111; DB 14; Length 1161;
Best Local Similarity 43.6%; Pred. No. 4.9e-10;
Matches 351; Conservative 0; Mismatches 454; Indels 0; Gaps 0;

```
Qy        498 GGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCGCTCGTGACGCTGCATCT 557
Db       1061 GGGAGAGGNGGANGGAGNNANGGGGAAGGGGGGNNGGGGNGGGGGGGGGGGGGGGGGNG 1002

Qy        558 GGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTGGCCGGCGGCGTCTCGGT 617
Db       1001 NGGGGGGGGGGGGGGGGGGGGGTTGNGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 942

Qy        618 CATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGCGGGCTGTCGGTGGACGG 677
Db        941 GGGGNNGGGGGGGGGGGGGGNNGGGGGNNGGGGGGGGGGGGGGGGGGGGNNNGGGGGGG 882

Qy        678 CAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGGGGCGAGGGCGTCGGGAT 737
Db        881 GNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNCCGGGGGGGGGGGGGNGGGNGGGGG 822

Qy        738 GCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCATCGGGTGCTGGCGGTGGT 797
Db        821 GGGGGGGGGGGGGGGGGNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 762

Qy        798 ACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTGACGGCGCCGAACGGTCC 857
Db        761 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGNGGGGGGGGGGGGGGGGGGGGGGG 702

Qy        858 GGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGGTTGTCCGTGGCGGATGT 917
Db        701 GGGGGGGGGGGGGGGGGGGGGGGGGGGNNGGGGGGGGGGGGNGGGGGGGGGGGGGGN 642

Qy        918 GGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGATCCGATCGAGGCACAGGC 977
Db        641 NGNNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 582

Qy        978 GTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTGTGGCTGGGGTCTCTGAA 1037
Db        581 GGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 522

Qy       1038 GTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGGGTCATCAAGATGGTGAT 1097
Db        521 GGGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGG 462

Qy       1098 GGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGATAAGCCGTCGCCGCAGGT 1157
Db        461 GGGGGGGGGGGGGGGGGNGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGNNGGGGGGG 402

Qy       1158 GGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTGCCGTGGCCGGGGGACGC 1217
Db        401 GGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGNGG 342

Qy       1218 GGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATCGGCGGCACGAATGCGCA 1277
Db        341 GGGGGGGGGGGGNNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGG 282
```

```
Qy       1278 TGTGATTTTGGAGGAGGCGCCGGCG 1302
Db        281 GGGGGGGGGGGGGGGNGGNGGGGGGG 257
```

RESULT 2
AG152606
LOCUS AG152606 1138 bp DNA linear GSS 09-JAN-2002
DEFINITION Pan troglodytes DNA, clone: RP43-016B03.TJ, genomic survey sequence.
ACCESSION AG152606
VERSION AG152606.1 GI:16682284
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee Male BAC Library clone:RP43-016B03.TJ.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library RPCI-43
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1138)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library RPCI-43 This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS
Sequencing: TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES Location/Qualifiers
source 1..1138
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="RP43-016B03.TJ"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT 38 a 9 c 869 g 38 t 184 others
ORIGIN

Query Match 8.5%; Score 110.6; DB 17; Length 1138;
Best Local Similarity 38.9%; Pred. No. 5.8e-10;
Matches 389; Conservative 0; Mismatches 610; Indels 0; Gaps 0;

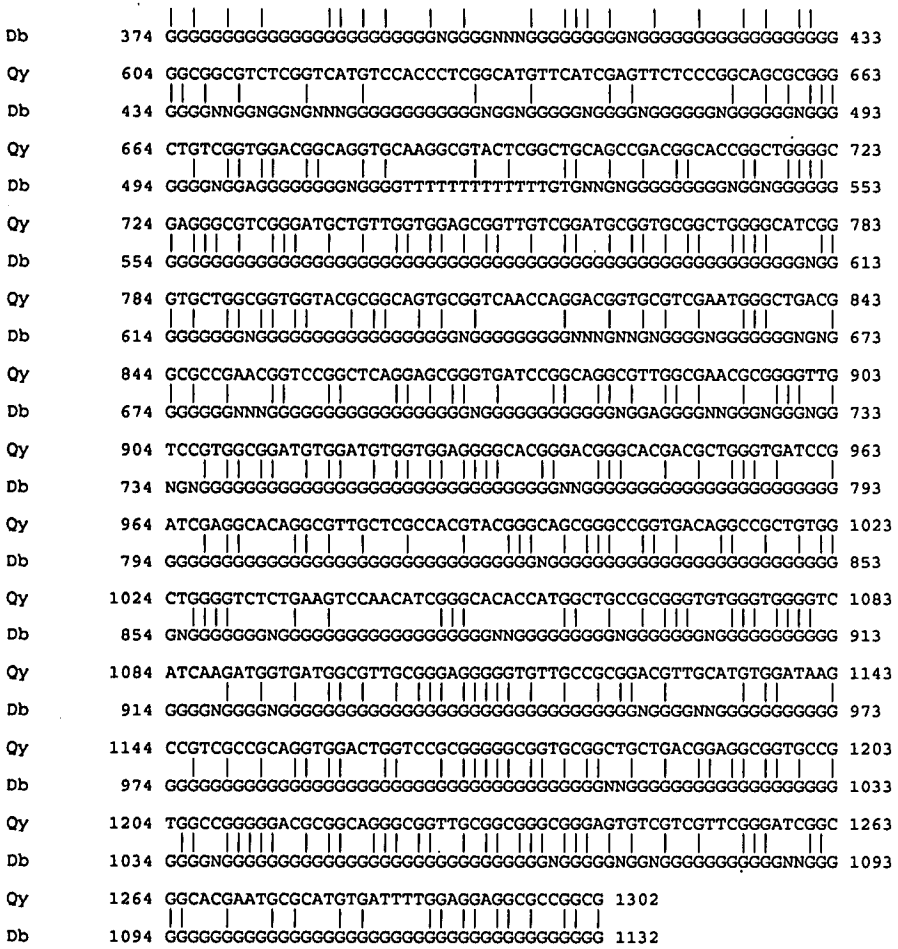
```
Qy        304 TTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTCGCG 363
Db        134 TTTGTGAGGGNCGTGGNNNNNNGCGTGGGNGTTGGGGGNNNNNNNNNNNAAAAANGAGGG 193

Qy        364 GGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCCGCG 423
Db        194 GGGGGGNNNNNNNNNNNNGNNNNNNNNGNNGGNGGGNNNGGGGGGGGGGGGGGGGGGG 253

Qy        424 GACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATCGCC 483
Db        254 GGGGGGGGGGGGNNGGNGGNNGGGNGNNNGGNNNGNNNGGGNGGGGGNGGGGGGGGGNN 313

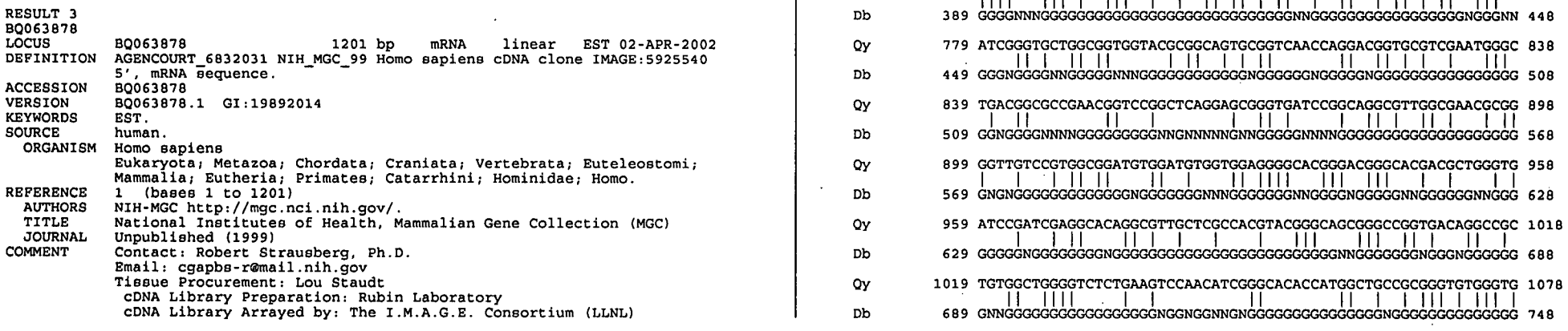
Qy        484 TACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCGCTC 543
Db        314 NGGGGGNNGGGGGGGGGNNNNNNGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGG 373

Qy        544 GTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTGGCC 603
```



RESULT 3
BQ063878
LOCUS BQ063878 1201 bp mRNA linear EST 02-APR-2002
DEFINITION AGENCOURT_6832031 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5925540 5', mRNA sequence.
ACCESSION BQ063878
VERSION BQ063878.1 GI:19892014
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2097 row: g column: 13
High quality sequence stop: 152.
FEATURES Location/Qualifiers
source 1. .1201
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5925540"
/clone_lib="NIH_MGC_99"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 23 a 41 c 763 g 15 t 359 others
ORIGIN

Query Match 8.5%; Score 110.2; DB 14; Length 1201;
Best Local Similarity 32.0%; Pred. No. 6.8e-10;
Matches 283; Conservative 0; Mismatches 601; Indels 0; Gaps 0;



```
Qy       1079 GGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGG 1138
              |||         |  ||  |  ||        |   ||||  |   |  ||      |   | ||
Db        749 GGGGGGGGGGGGGGGGGGNGGGGGGGNGNGNNNGGGGGGGGGGGGGGGGGGNNGGGGGGGGG 808

Qy       1139 ATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGG 1198
                |  |  |  |  || |     |    |  |||| || | ||  |  |    | || ||
Db        809 GGGGGGGGGGGGGGGGGGGGGNGNNGNNGGGGNGGGGGGGGGGGGGGGGGGGNNGGGGGGG 868

Qy       1199 TGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGA 1258
                |  |  |||||          ||| ||  | || ||| ||| | |  |  |      |
Db        869 GGGGGGGNGGGGGGGGNGNGNNGGGGGNGGGNGNGGGGGGGGGGGGGGGGGGNGNNGNNGG 928

Qy       1259 TCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
                || ||       | |   | |       |  ||||   | |
Db        929 GGGGGGGGGNNGGGGGGGGGGGGGGNGGNNGGNNGGCGGNNGGG 972
```

RESULT 4
BM913931
LOCUS BM913931 1243 bp mRNA linear EST 12-MAR-2002
DEFINITION AGENCOURT_6608063 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5478793 5', mRNA sequence.
ACCESSION BM913931
VERSION BM913931.1 GI:19364310
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1243)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1999 row: a column: 02
High quality sequence stop: 369.
FEATURES Location/Qualifiers
source 1. .1243
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5478793"
/clone_lib="NIH_MGC_98"
/tissue_type="astrocytoma grade IV, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 47 a 193 c 776 g 116 t 111 others
ORIGIN

Query Match 8.4%; Score 109.4; DB 14; Length 1243;
Best Local Similarity 41.3%; Pred. No. 9.3e-10;
Matches 353; Conservative 0; Mismatches 502; Indels 0; Gaps 0;

```
Qy        411 GGAGGGCGCCGCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTC 470
              ||  || | || ||    |  || ||  |  ||   | ||       |||    |  |||
Db        185 GGGCGGTGGCGTGGGATCGCGGGTCCTTACGCTGGGCTGTCTGGTCAGCACGTGCAGGTC 244

Qy        471 GGGCCGTATCGCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTG 530
```

```
Db        245 AGGGCAGGTCCTCTGAGCCGGCGCCCCTGGCCAGCAGGCGAGGCTACAGTACCTGCTGTC 304
              || |   ||  ||   |  ||  |  |   || |||||       | ||  |  ||
Qy        531 CTCGGCATCGCTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCAC 590
              |       |  | | | |  |||  ||    |  ||  |  |  ||||||  |
Db        305 TTTCCAGGGGGAAGGGGCTCCCCATGAGGGAGGGGCGACGGGGGAGGGGGGTGATGGNGN 364

Qy        591 GCTCGCGCTGGCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTC 650
                  |  | |  |   ||                  |     | |
Db        365 NNGNNNNNNNGNGNGNGNNGNNGGGGGNNGNGGNGNNGNNNNGNNGGGGGGNGGGGNNGN 424

Qy        651 CCGGCAGCGCGGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGG 710
               ||  | | |||  |    |   ||  || |  || |     ||  |  |  | ||
Db        425 NGGGGGGGGGGGGGGGGGNGGNNGGGGNGNGGNGGGGGGGGGGGGGGGGGGGNGGNGGG 484

Qy        711 CACCGGCTGGGGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCG 770
               ||  |||| | |||  |  |||  |  |  || || | ||  |  ||  | |  | ||
Db        485 GGNGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGNGGGGGGGGGGGGGGNNGN 544

Qy        771 GCTGGGGCATCGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTC 830
               ||||   ||| |  || || ||   | ||  | | ||        ||  || | |
Db        545 NGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGNGGGGGGG 604

Qy        831 GAATGGGCTGACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGC 890
               |  |||  |  |  |  |  ||  |  |  |||  |  ||  || |  |
Db        605 GGGGGGGGGGGNGNGGGGGGGGGGGGNGNGGNGNNNNGGGGGGGGGGGGGGGGGGNGT 664

Qy        891 GAACGCGGGGTTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGAC 950
               | ||||  |  | || ||  | || || | || || |||| | |||  |||  |
Db        665 TTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 724

Qy        951 GCTGGGTGATCCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGA 1010
               | |||  |   |  |||   || |      |   |  |||  | |||  || |
Db        725 GGGGGGGGGGGGGGGGGGGGNGGGGGGGGNNGGGGGGGGGGGGGGGGGGGGGGGGGG 784

Qy       1011 CAGGCCGCTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGG 1070
               ||  |  | ||  ||||    |  |           ||      ||  |  | |||
Db        785 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGNGGGGGGGGGGGGGNGGGGGG 844

Qy       1071 TGTGGGTGGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTT 1130
               | ||| ||||        |  || |  || |  | ||| ||||| |  |  |  |  |
Db        845 GGGGGGGGGGGNNGGNGNGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGNGGGG 904

Qy       1131 GCATGTGGATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGAC 1190
               |   | ||   |  |  |  |  |  ||   ||   | ||||   | |  |  |  |
Db        905 GGGGGGGGGGGGGGGGGNGGGGGNGNGGGGGGGGGGGGGGGGGNGNGGGGNGNNGNNGGN 964

Qy       1191 GGAGGCGGTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTC 1250
               || || || |   ||  |||||  | ||  |||  ||| | |   ||  ||   |  |
Db        965 GGGGGGGGGGGGGNNGGGGGGGGGGGGGGGGGGGGGGGGTGGGGNNNGGGTGGNNNGGGGGG 1024

Qy       1251 GTTCGGGATCGGCGG 1265
               |   |||   || ||
Db       1025 GGGGGGGGGGGGNGG 1039

RESULT 5
BI416575
LOCUS BI416575 931 bp mRNA linear EST 15-AUG-2001
DEFINITION hasp001xg15f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xg15f, mRNA sequence.
ACCESSION BI416575
VERSION BI416575.1 GI:15187598
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 931)
AUTHORS Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
```

```
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Fred O. Asiegbu
            Dept. of Forest Mycology & Pathology
            Swedish University of Agriculture, Box 7026,S-750 07, Uppsala,
            Sweden
            Tel: +46 18 67 15 98
            Fax: +46 18 30 92 45
            Email: Fred.Asiegbu@mykopat.slu.se
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .931
                     /organism="Pinus sylvestris/Heterobasidion annosum"
                     /db_xref="taxon:169015"
                     /clone="hasp001xg15f"
                     /clone_lib="Heterobasidion annosum - Scots pine infection
                     stage subtraction cDNA library (hasp)"
                     /dev_stage="Seedling roots of scots pine were infected for
                     6 days with H. annosum"
                     /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive
                     hybridization cDNA library was constructed from  scots
                     pine roots infected for 6-days with mycelia of
                     Heterobasidon  annosum (PP5)."
BASE COUNT        9 a      7 c    884 g     15 t     16 others
ORIGIN

  Query Match             8.3%;  Score 108;  DB 13;  Length 931;
  Best Local Similarity   47.5%;  Pred. No. 1.6e-09;
  Matches  306;  Conservative    0;  Mismatches  338;  Indels    0;  Gaps    0;

Qy        653 GGCAGCGCGGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCA 712
              ||  | | |||  |  || ||  ||  || |  || |    ||  |  |  |  ||
Db        243 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGG 302

Qy        713 CCGGCTGGGGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGC 772
              ||  |||| | ||| |  |||  |  |  |  || | ||  |  ||  | || | ||
Db        303 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGG 362

Qy        773 TGGGGCATCGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGA 832
              ||||    ||| |  || || ||   | ||  | | ||        ||  || | |  |
Db        363 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 422

Qy        833 ATGGGCTGACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGA 892
              |||  |  || |  |  ||  ||   || | ||| |    ||  || |  || |
Db        423 GGGGGGGGGGGGGGGGGGGGGNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 482

Qy        893 ACGCGGGGTTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGC 952
              | ||||  |   | || ||  | ||  | || || ||||   |||  |||   |  |
Db        483 GGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGT 542

Qy        953 TGGGTGATCCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACA 1012
              |||  |    |  | |||   || |  |   |    |  |||  | |||  || |
Db        543 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 602

Qy       1013 GGCCGCTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTG 1072
              ||  |  | ||  ||||   ||  |    |||   |  ||  |  | |||||
Db        603 GGGGGGGGGGGGGGGGGGGGGGTGGGGGGGGGGGGGGGGGGGTGGGGGGGGGGGGGGGTG 662

Qy       1073 TGGGTGGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGC 1132
              |||  ||||        |  || |  || |  ||| |||||||  |  | ||  | |
Db        663 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGTGGGGGGGGGGGGGGGG 722

Qy       1133 ATGTGGATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGG 1192
              | ||    |  |  |  | || ||   ||   | ||||| || | ||  |  |  ||
Db        723 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 782

Qy       1193 AGGCGGTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGT 1252
              || ||||  | ||  |||||  | ||  |||  |  | || ||| ||| | |  |  |
Db        783 GGGGGGTGGGGGGGGGGGGGGGNGGGGGGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGG 842

Qy       1253 TCGGGATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCG 1296
              |||   || ||   |  |  | |  | | || || || |
Db        843 GNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGTGGGGGGGGGGG 886
```

```
RESULT 6
AG146864
LOCUS       AG146864               1037 bp    DNA     linear   GSS 08-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-008J12.TJ, genomic survey
            sequence.
ACCESSION   AG146864
VERSION     AG146864.1  GI:16676542
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
            Male BAC Library clone:RP43-008J12.TJ.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1037)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: TJ
             LIBRARY
                 Vector    : pBACe3.6
                 R.Site 1  : EcoRI
                 R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .1037
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-008J12.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT       64 a     35 c    765 g     26 t    147 others
ORIGIN

  Query Match             8.3%;  Score 108;  DB 17;  Length 1037;
  Best Local Similarity   42.6%;  Pred. No. 1.6e-09;
  Matches  282;  Conservative    0;  Mismatches  380;  Indels    0;  Gaps    0;

Qy        604 GGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGCGGG 663
              | |||                                          ||  | | |||
Db        133 GCCGGNNNNNNNNNNNNNNNNNNNNNNNNNGGGNGGNNNNNNNGGNNGNGGNNGGGGGG 192

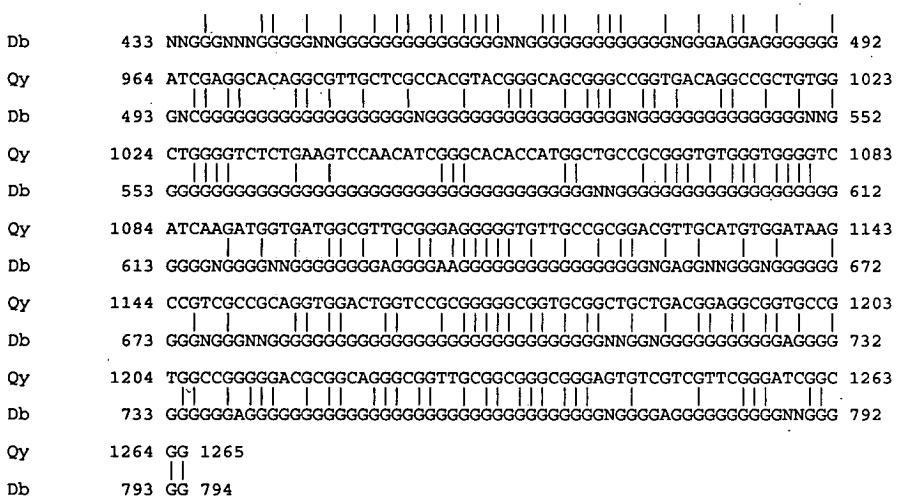
Qy        664 CTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGGGGC 723
              |  || ||  ||  || |   || |    ||  |  |  |  |    ||  ||||
Db        193 GGGGGGGGGGGGGGGGGGGGGGGNGGGGNNGGGGGGGGGGGGGGGGGNGGGGGGGGGGG 252

Qy        724 GAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCATCGG 783
              | ||| |  |||  |  || || | ||  |  ||  | | ||  |  |||     ||
Db        253 GGGGGGGGGGGGGGGNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 312

Qy        784 GTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTGACG 843
              | |    || ||   | ||  | | ||        ||  || | |  |   ||| | ||
Db        313 GGGGGNNGGGGGGGNGGGGGGGGNGGGGGNGNNGGGGGGGGGGGGGGGGGGGGGGGCG 372

Qy        844 GCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGGTTG 903
              | ||   |  | || |     | ||| |    || |||    || |  | |||| |
Db        373 CCCCCCCTCTTTGTGGTTTTCNNGNGGGGGGGGAGGGAGGNNNAGGGGGGGGGGGGGGG 432

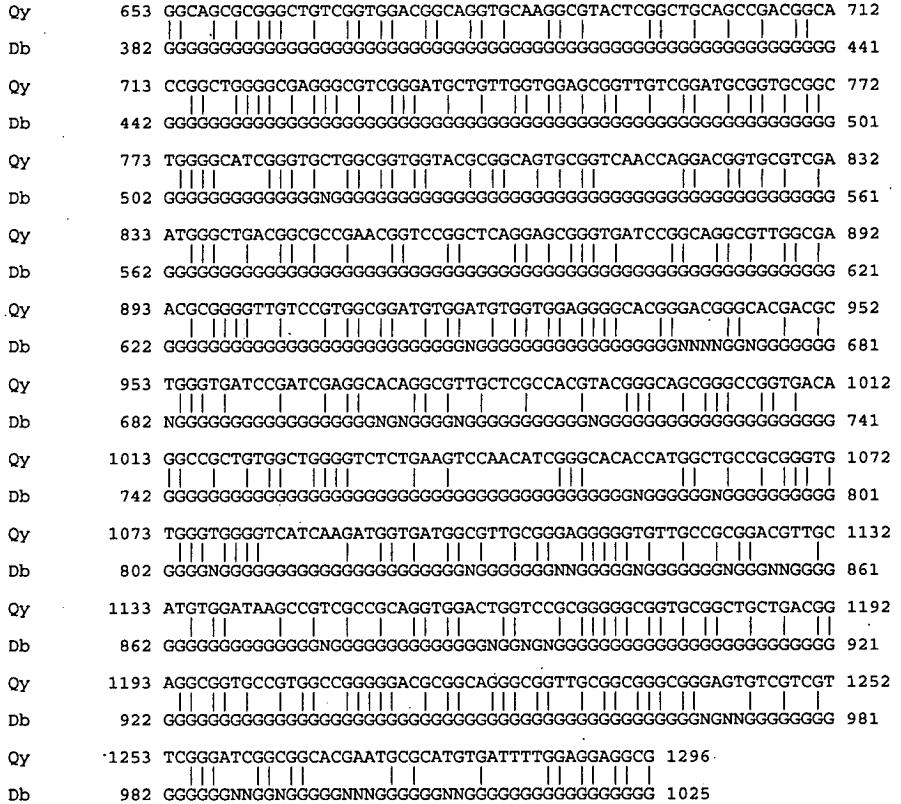
Qy        904 TCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGATCCG 963
```



RESULT 7
BQ930965
LOCUS BQ930965 1089 bp mRNA linear EST 20-AUG-2002
DEFINITION AGENCOURT_8931577 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6483877 5', mRNA sequence.
ACCESSION BQ930965
VERSION BQ930965.1 GI:22345996
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1089)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2669 row: c column: 14
High quality sequence start: 28
High quality sequence stop: 180.
FEATURES Location/Qualifiers
source 1. .1089
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6483877"
/clone_lib="NIH_MGC_40"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 26 a 54 c 746 g 30 t 233 others

ORIGIN

Query Match 8.3%; Score 107.8; DB 14; Length 1089;
Best Local Similarity 45.8%; Pred. No. 1.8e-09;
Matches 295; Conservative 0; Mismatches 349; Indels 0; Gaps 0;



RESULT 8
BM809948
LOCUS BM809948 1288 bp mRNA linear EST 05-MAR-2002
DEFINITION AGENCOURT_6581186 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5454872 5', mRNA sequence.
ACCESSION BM809948
VERSION BM809948.1 GI:19126771
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1288)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation

```
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM1950  row: 1  column: 09
          High quality sequence start: 7
          High quality sequence stop: 357.
FEATURES             Location/Qualifiers
     source          1. .1288
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5454872"
                     /clone_lib="NIH_MGC_98"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT       75 a    131 c    783 g     62 t    237 others
ORIGIN

  Query Match             8.3%;  Score 107.8;  DB 14;  Length 1288;
  Best Local Similarity   39.4%;  Pred. No. 1.8e-09;
  Matches  256;  Conservative    0;  Mismatches  393;  Indels    0;  Gaps    0;

Qy        654 GCAGCGCGGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCAC 713
              |   |   |  ||     ||     | |   ||   |   |   |   |
Db        414 GGNNNGNNNGNNNGNGGNNNNNNNNNGGNNNNNGGGGGGGNGGGNGNGGNNGGNNNGGN 473

Qy        714 CGGCTGGGGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCT 773
              ||  ||||    || |  |||  |       | | ||  |  ||  | || | |
Db        474 GGGGGGGGGNNNNGGNGGGGGGNNGNNNGNNNNNGGGGGGGGGGGGGGGGGGGGNNGNG 533

Qy        774 GGGGCATCGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAA 833
              ||||    ||| |  || |  ||  | ||  | | |        ||  ||   | |
Db        534 GGGGGNGNGGGGGGGGGGNNGGGGGGGGGGGGGGGGNNNGGNGGGGNGGGNGGGNGGG 593

Qy        834 TGGGCTGACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAA 893
              |||  |  || |  |    |  ||   || | |||       ||   | |  || |
Db        594 GGGGGGGNGGGGGGGGGNNNGNGGGGNNNGGGGGGGGGGNNNGGNGGNNNGGGGGGGGG 653

Qy        894 CGCGGGGTTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCT 953
              |   ||  |   | || ||                | || || ||||   |||  | |   |
Db        654 GGGNNGGGGGNNNGNGGGGGGNNGNNGGGGGGGGGGGGGGGGGGGGGNGNGGGGGGNNGG 713

Qy        954 GGGTGATCCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAG 1013
              ||| |   |  | ||   || | |  |   |   |  || | ||  |       |
Db        714 GGGGGGGNNGGGGGNGGGGNNGGGGGGGGGGGGGGGGNGGNGGGGGGNGGGGGNGNGGNG 773

Qy       1014 GCCGCTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGT 1073
              |   |  | ||  |||     |  |       | |     || |  | | ||  |
Db        774 GGGGNNGGGGGGNGGGGGGNNGGGGGGNNNGNGGGNGGGGGGGGGGGGGGGGGGNNGG 833

Qy       1074 GGGTGGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCA 1133
              |||  ||||        | || |  | | |  |||  ||||  |   |  || | |
Db        834 GGGGGGGGGGGGGGGGNGGGNGGNGGNGNNGGGGGGNGGGGGGGNGNGNNGGGNGGGGGG 893

Qy       1134 TGTGGATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGA 1193
              | ||    |  |  |  |  |  ||  ||   | ||||| || | ||      | ||
Db        894 GGNGGGGGGGGGGGGGNGGNGGNNGGGGGGGGGGGGGGGGGGGGGGGGGNGNGGGGGG 953

Qy       1194 GGCGGTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTT 1253
              || || |  | ||  ||  |  |  | ||  ||   || ||| ||| | |  |  |
Db        954 GGGGGGGNGGGGGGNGGNNGGGGGNGGGGGNNGGGGNGGGGGGGGGGGGGGGGGGGGGG 1013

Qy       1254 CGGGATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              ||    |||||  |  |  |  |  || || || |  ||| |
Db       1014 NGGNGGGGGCGGNGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1062
```

```
RESULT 9
AZ189327
LOCUS       AZ189327                888 bp    DNA     linear   GSS 30-AUG-2000
DEFINITION  SP_1014_A1_F05_T7A Strongylocentrotus purpuratus, purple sea urchin
            , sperm genomic BAC library Strongylocentrotus purpuratus genomic
            clone Plate=1014 Col=9 Row=K, DNA sequence.
ACCESSION   AZ189327
VERSION     AZ189327.1  GI:8372605
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 888)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and
            Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 1014  row: K  column: 9
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 888.
FEATURES             Location/Qualifiers
     source          1. .888
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="Plate=1014 Col=9 Row=K"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
BASE COUNT       52 a     37 c    606 g     83 t    110 others
ORIGIN

  Query Match             8.2%;  Score 107.2;  DB 17;  Length 888;
  Best Local Similarity   44.4%;  Pred. No. 2.3e-09;
  Matches  274;  Conservative    0;  Mismatches  343;  Indels    0;  Gaps    0;

Qy        653 GGCAGCGCGGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCA 712
              ||  | |       |  || ||  ||  || |   ||        ||  |  |  |  ||
Db        272 GGGGGNGNNNNNNGGGGGNGGGGGGGGGGNGGGGGGGNGGGGGGGGGGGGGGGGGGGGG 331

Qy        713 CCGGCTGGGGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGC 772
                ||  |||| | ||| |    |  |  |  || || | ||  |  ||  | || | ||
Db        332 GGGGGGGGGGGGGGGGGGGGNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 391

Qy        773 TGGGGCATCGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGA 832
              ||||    ||| |  || || ||   | ||  | | ||        ||  || | |  |
Db        392 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNNGGNGGGGGGGGGGGG 451

Qy        833 ATGGGCTGACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGA 892
              |||  |  || |  |    |  ||   || | |||       |   | |  || |
Db        452 GNGGGGGGGGGGGGGGGNNGNGNNGNGGGGGGGGGGGGGGGGGGGGGNGGNGGGGGGGGGG 511

Qy        893 ACGCGGGGTTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGC 952
              | ||||  |   | || ||  | ||     || || ||||   |||  |||   |  |
Db        512 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNGGGGGNGGGGGGGGGGGGGGGGGGGGG 571

Qy        953 TGGGTGATCCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACA 1012
              ||        |   | ||    || |  |   |     |  |||  | |||  || |
```

```
Db        572 GGGNGNNNGNGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGGNGGGGGGGN 631

Qy       1013 GGCCGCTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTG 1072
              ||  |  | ||  ||||    |  |          |||        ||  |  | ||| |
Db        632 GGGGGGGGGGGGGGGGGGGGGNGGGGNGGGGGGGGGGGNGGGGGGGGGNGGNGGGGGGGG 691

Qy       1073 TGGGTGGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGC 1132
              ||| ||||        |  || |  |  |  | ||| ||||| |  |  | ||  |  |
Db        692 GGGGGGGGGGGGGGGGGGGGGNGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGG 751

Qy       1133 ATGTGGATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGG 1192
                | ||    |  |  |    |  |    ||   | ||||| || | ||  |    ||
Db        752 NNGGGGGNGGGNGGGGGNGNNNGNNGNNGGGGNNGGGGGGGGGGGGGGGGGGGNGNNGGG 811

Qy       1193 AGGCGGTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGT 1252
              ||  | |  | ||  |||||  | ||  ||| ||  | |  ||| ||| | |  |  |
Db        812 GGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNGGGGGGGGGGGGGGGG 871

Qy       1253 TCGGGATCGGCGGCACG 1269
                |||   || ||  ||
Db        872 GGGGGGGGGGGGGCCG 888


RESULT 10
AQ895329
LOCUS       AQ895329                 918 bp    DNA     linear   GSS 10-NOV-1999
DEFINITION  HS_4832_A2_H01_T7A CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=4832 Col=2 Row=O, DNA sequence.
ACCESSION   AQ895329
VERSION     AQ895329.1  GI:6351435
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 918)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 4832  row: O  column: 2
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 918.
FEATURES             Location/Qualifiers
     source          1. .918
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=4832 Col=2 Row=O"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT       16 a     93 c    718 g     60 t     31 others
ORIGIN

  Query Match             8.2%;  Score 107.2;  DB 17;  Length 918;
  Best Local Similarity   49.1%;  Pred. No. 2.3e-09;
  Matches  301;  Conservative    0;  Mismatches  307;  Indels    5;  Gaps    1;

Qy        653 GGCAGCGCGGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCA 712

              ||  | ||| |  |  || ||  ||  || |   || |      | |    || |  ||
Db        305 GGGGGGGCGCGGCGGGGGCGGNGGGGGGGGGGGGGGGGGGGGCCGCGGGCGGGGGGGG 364

Qy        713 CCGGCTGGGGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGC 772
              |||  | || | |||  |  ||||  |   |  || || | ||  |  ||  | || | ||
Db        365 GCGGGGGCGGGGGGGGGGGCGGGCGGGGGGGGGGGGGGGGGGGGGGCNGGGGGGGGGG 424

Qy        773 TGGGGCATCGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGA 832
              ||||    ||  |  || || ||  |  ||  | | ||          ||  || | |  |
Db        425 GGGGGGGGGGNGGGGGGGGGGGGGCCCGGGGGGGGGGGGGGCGGGGGGGGGGGGGGGG 484

Qy        833 ATGGGCTGACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGA 892
              |||  |  || |  |   ||   ||   || | ||| |  |||||  | |  || |
Db        485 GGGGGGGGGGGGNGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGCCGGCGCGGGGGGGGC 544

Qy        893 ACGCGGGGTTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGC 952
              | |||| |  ||  | ||  | ||  |  | |  |||||  |||  |||   | |
Db        545 GGGGGGGGGGGGGCGGNGGGGGGGGGGGGGGGCGGGCGGGGGCGGGGGGGGGGGNGCGGG 604

Qy        953 TGGGTGATCCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACA 1012
              ||| |    |  || ||    || |  |  | |  |   ||| | |||  ||
Db        605 GGGGGGCCGGGGGCGCGGGGGGGGGGCGGGGGGCGGGGGGGGGGGGGGGGGGCGGC 664

Qy       1013 GGCCGCTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTG 1072
              |||||  | ||| ||||  |   |  | |       |||        ||     | ||| |
Db        665 GGCCGGGGGGGCGGGGGGCGGGGCGGC-----GGGGGGGGGGGGGGGGGCGGGGGGGCG 719

Qy       1073 TGGGTGGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGC 1132
              |||  | | |  |  | || |  ||  |  | ||  ||||  |  || | ||  |  | |
Db        720 GGGGGCGCGGCGGCGGGCCGGGGGGGGGGGGGGGGCGGGGGGGGGGGCNGGGCGGCCGG 779

Qy       1133 ATGTGGATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGG 1192
              | ||    |  |  |  |  || ||   ||   | |||||  || ||||  |  |  ||
Db        780 GGGGGGGGGGGGGGGGGGGGCGGGGGGGGGGGGNGGGGGGGGGGCGCGGGGGGGGCGGG 839

Qy       1193 AGGCGGTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGT 1252
              || || |  |  || |||||  || |||  |||  ||   | ||  |||||||  | |  ||
Db        840 GGGGGGGGGGGGCGCGGGGGGGCGGGGCGGGGGGGGGGGGGGGGGCGGGGGGGCGGGCG 899

Qy       1253 TCGGGATCGGCGG 1265
                |  |   |||  ||
Db        900 GGGCGGGCGGGGG 912


RESULT 11
AI514012
LOCUS       AI514012                 662 bp    mRNA    linear   EST 23-APR-2001
DEFINITION  GH27052.5prime GH Drosophila melanogaster head pOT2 Drosophila
            melanogaster cDNA clone GH27052 5 similar to BcDNA:GH07626:
            FBan0003523 located on: 2L 23D1-23D1;: 04/10/2001, mRNA sequence.
ACCESSION   AI514012
VERSION     AI514012.1  GI:4418074
KEYWORDS    EST.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 662)
  AUTHORS   Harvey,D., Brokstein,P., Hong,L., Evans-Holm,M., Su,C., Tsang,G.,
            Lewis,S. and Rubin,G.M.
  TITLE     BDGP/HHMI Drosophila EST Project
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: GH27052.3prime
            Contact: Stapleton, M.
            BDGP
            Lawrence Berkeley National Lab
            One Cyclotron Rd, Berkeley, CA 94720, USA
            Fax: 510 486 6798
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            hit genomic AE003581: arm:2L [2747932,3054367]
            estimated-cyto:23B3-23D3: 04/10/2001
            Plate: GH.270  row: E  column: 4
```

```
          High quality sequence stop: 542
          POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .662
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="GH27052"
                     /clone_lib="GH Drosophila melanogaster head pOT2"
                     /sex="male and female"
                     /dev_stage="adult"
                     /lab_host="DH5 - alpha"
                     /note="Organ: head; Vector: pOT2; Site_1: EcoRI; Site_2:
                     XhoI; Sized fractionated cDNAs were directly ligated into
                     pOT2.  Plasmid cDNA library."
BASE COUNT      143 a    185 c    185 g    149 t
ORIGIN

  Query Match             8.2%;  Score 107;  DB 9;  Length 662;
  Best Local Similarity   50.7%;  Pred. No. 2.5e-09;
  Matches  316;  Conservative    0;  Mismatches  295;  Indels   12;  Gaps    2;

Qy        529 TGCTCGGCATCGCTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGC 588
              ||||| || ||     | || | | ||| | |       ||||   ||  ||  |
Db          3 TGCTCCAGTTCTCTGTACGCCTTGGAACAGGCTTTCTCCGATATGCGCGAAGGAAAGGTC 62

Qy        589 ACGCTCGCGCTGGCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTC 648
              ||| |||| ||  || |       | ||   ||  | |  |||||    |  |||||
Db         63 GACAACGCCCTGGTCGCTGGAGCTGGACTGATCCTCAAGCCCACCATGTCGCTGCAGTTC 122

Qy        649 TCCCGGCAGCGCGGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGAC 708
              || | |   | ||     |||||||| |||||||| | |   |  | | | |
Db        123 AAGCGACTGAATATGTTGAGCCCGGACGGCAGCTGCAAGGCCTTCGATGAGTCTGGCAAT 182

Qy        709 GGCACCGGCTGGGGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTG 768
              ||  || | |   ||| ||    | | ||||| |  || ||||    || |  ||   |
Db        183 GGATACGTCCGTTCCGATGGATGTGTGGTGCTGCTCCTGCAGCGCACCTCTGCAGCCAGG 242

Qy        769 CGGCTGGGGCATCGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCG 828
              ||  ||          | ||    |||      |  | |  |    | ||||
Db        243 CGTGTGTATGCCTCCATCCTCAATGTGCGCACCAATACGGATGGTTTCAAGGA------- 295

Qy        829 TCGAATGGGCTGACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTG 888
              | | ||  | ||    || |  ||   |   ||  | ||  |||||||  || | |
Db        296 --GCAGGGCATCACATACCCTATTGGCAAGATGCAAAATCGCCTGATCCGCGAGACCTAC 353

Qy        889 GCGAACGCGGGGTTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACG 948
              | | |    ||  |   | || |||||||  |    |||||||  ||||| || || ||
Db        354 GAGGAGATTGGTCTTAACCCCGCCGATGTGGTTTACGTGGAGGCACACGGTACCGGAACC 413

Qy        949 ACGCTGGGTGATCCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGT 1008
              | | |||| |||||     ||||   |  | |    | ||  | | |  |   || | ||
Db        414 AAGGTGGGCGATCCCCAGGAGGTGAACTCTATCACTGACTTCTTCTGCAA---GGACCGT 470

Qy       1009 GACAGGCCGCTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCG 1068
                |  || |||| | | || ||  | ||||||||||| || |||  |     | ||  |
Db        471 ACGACCCCTCTGTTGATCGGATCGGTCAAGTCCAACATGGGTCACTCGGAACCCGCTTCC 530

Qy       1069 GGTGTGGGTGGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACG 1128
              |||||  |    ||  ||||||  |||| ||  ||  |||||| ||  | || |  |
Db        531 GGTGTCTGCTCTGTGGCCAAGATTCTGATCGCCATGGAGGAGGGCGTCATTCCCGGTAAC 590

Qy       1129 TTGCATGTGGATAAGCCGTCGCC 1151
              |||||     | ||||||   ||
Db        591 TTGCACTACAACAAGCCGAACCC 613


RESULT 12
BI416699
LOCUS       BI416699                 863 bp    mRNA    linear   EST 15-AUG-2001
DEFINITION  hasp001xp01f Heterobasidion annosum - Scots pine infection stage
            subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
            annosum cDNA clone hasp001xp01f, mRNA sequence.
ACCESSION   BI416699
VERSION     BI416699.1  GI:15187722
KEYWORDS    EST.
SOURCE      Pinus sylvestris/Heterobasidion annosum.
  ORGANISM  Pinus sylvestris/Heterobasidion annosum
            Eukaryota; mixed EST libraries.
REFERENCE   1  (bases 1 to 863)
  AUTHORS   Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
  TITLE     Expressed sequence tags of randomly selected cDNA clones from the
            interaction of the root rot fungus (Heterobasidion annosum) with
            seedling roots of Scots pine (Pinus sylvestris)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Fred O. Asiegbu
            Dept. of Forest Mycology & Pathology
            Swedish University of Agriculture, Box 7026,S-750 07, Uppsala,
            Sweden
            Tel: +46 18 67 15 98
            Fax: +46 18 30 92 45
            Email: Fred.Asiegbu@mykopat.slu.se
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .863
                     /organism="Pinus sylvestris/Heterobasidion annosum"
                     /db_xref="taxon:169015"
                     /clone="hasp001xp01f"
                     /clone_lib="Heterobasidion annosum - Scots pine infection
                     stage subtraction cDNA library (hasp)"
                     /dev_stage="Seedling roots of scots pine were infected for
                     6 days with H. annosum"
                     /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive
                     hybridization cDNA library was constructed from  scots
                     pine roots infected for 6-days with mycelia of
                     Heterobasidon  annosum (FP5)."
BASE COUNT       12 a     21 c    700 g     12 t    118 others
ORIGIN

  Query Match             8.2%;  Score 106.8;  DB 13;  Length 863;
  Best Local Similarity   40.4%;  Pred. No. 2.7e-09;
  Matches  327;  Conservative    0;  Mismatches  482;  Indels    0;  Gaps    0;

Qy        494 GGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCGCTCGTGACGCTGC 553
              |||   |||||  || |       |  |       ||   |||        | |
Db          1 GGCGACAAGGGGGGGGGGGGGGGGGCGGGNGNNGNGCNNGGGTTNNGCTACNTGAACAAC 60

Qy        554 ATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTGGCCGGCGGCGTCT 613
              || |||    |  | |    ||||   ||      |   |       |
Db         61 ATNTGGNNNNCGGGGCTGGTCGGTAACGTCCCCCCNNNNCGNNGNNGNNNNTACGNNNNN 120

Qy        614 CGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGCGGGCTGTCGGTGG 673
              |     |                    |       |  |||||          ||
Db        121 NGNNNNNNTGNNNNNNNNNGNNNNNNNNGGGNNNGGGGGNGGGNGGCGGGGNNNNNNGG 180

Qy        674 ACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGGGGCGAGGGCGTCG 733
              ||  || |   || |       ||  |  |  |  ||    ||  |||| | ||| |  |
Db        181 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 240

Qy        734 GGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCATCGGGTGCTGGCGG 793
              ||  |  |  || || | ||  |  ||  | || |       |||    ||| |  || ||
Db        241 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNNGNNGGGGGGGGGGNGGGGGNGG 300

Qy        794 TGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTGACGGCGCCGAACG 853
              ||   | ||  | | ||       ||  || |    |  |||  |  || |      |
Db        301 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGNGGGGGGGGGGGGGGNGGGG 360

Qy        854 GTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGGTTGTCCGTGGCGG 913
              |   ||      | | ||  |   ||  | | |  | ||||  |    | || ||
Db        361 GGGGGGGGGGNGGGGGGNNGGNNGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 420

Qy        914 ATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGATCCGATCGAGGCAC 973
              | ||  | || || ||||   |||  |||   |  |  |||  |    |  | |
Db        421 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGG 480

Qy        974 AGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTGTGGCTGGGGTCTC 1033
              || | |   |   |  |||  | |||  || |   ||  |  | ||  ||||
```

```
Db        481 GGGGGGGGGGNGNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGG 540

Qy       1034 TGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGGGTCATCAAGATGG 1093

Db        541 GGGGGGGGGGGNNGGGGGGNNGGNGGGGNGGGGNGGGGGGGGGGGGGGGGGGGGGGGGG 600

Qy       1094 TGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGATAAGCCGTCGCCGC 1153

Db        601 GGGGGGGGNGGGGGGNNGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 660

Qy       1154 AGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTGCCGTGGCCGGGGG 1213

Db        661 GNGGGGGGNNNNGGNNGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGG 720

Qy       1214 ACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATCGGCGGCACGAATG 1273

Db        721 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGNGGGGGGGGGGGGNNNGGGNN 780

Qy       1274 CGCATGTGATTTTGGAGGAGGCGCCGGCG 1302

Db        781 GGGGGGGGGGGGGGGGGGGGGGGGGNGG 809
```

RESULT 13
BI416650
LOCUS BI416650 1015 bp mRNA linear EST 15-AUG-2001
DEFINITION hasp001xm01f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xm01f, mRNA sequence.
ACCESSION BI416650
VERSION BI416650.1 GI:15187673
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 1015)
AUTHORS Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026,S-750 07, Uppsala, Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES Location/Qualifiers
source 1. .1015
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="hasp001xm01f"
/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
/note="Vector: pT-Adv; Site_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidon annosum (FP5)."
BASE COUNT 21 a 6 c 961 g 16 t 11 others
ORIGIN

Query Match 8.1%; Score 105.8; DB 13; Length 1015;
Best Local Similarity 47.4%; Pred. No. 4e-09;
Matches 308; Conservative 0; Mismatches 342; Indels 0; Gaps 0;

```
Qy        653 GGCAGCGCGGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCA 712

Db        178 GGGAGGGGGGGGGGNGGGGGNGGGGTGGGGGGGGGGGGNGGGGGGCGGGGGGGGGGGGG 237

Qy        713 CCGGCTGGGGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGC 772

Db        238 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGAGGGGAGGGGGGNGGGGGGGGGG 297

Qy        773 TGGGGCATCGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGA 832

Db        298 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGA 357

Qy        833 ATGGGCTGACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGA 892

Db        358 GGGGGGGGGGGGGGGGGGGGTGGGGGGGGGGGGGGGGGGGAGGGGGGGGGGGGGGGGGG 417

Qy        893 ACGCGGGGTTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGC 952

Db        418 AGGGGGGGGGGGGGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 477

Qy        953 TGGGTGATCCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACA 1012

Db        478 GGTGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGTGGGGGGGGGGTAGG 537

Qy       1013 GGCCGCTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTG 1072

Db        538 GGGGGGGGGGGGGGGGGGGGGGGTGGGAGGGGGGGGGGGGGGGGGGGAGGGGGGGGGG 597

Qy       1073 TGGGTGGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGC 1132

Db        598 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 657

Qy       1133 ATGTGGATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGG 1192

Db        658 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGTGGGGG 717

Qy       1193 AGGCGGTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGT 1252

Db        718 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 777

Qy       1253 TCGGGATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302

Db        778 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 827
```

RESULT 14
BQ939440
LOCUS BQ939440 1089 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT_8929085 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6484345 5', mRNA sequence.
ACCESSION BQ939440
VERSION BQ939440.1 GI:22354918
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1089)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2670 row: g column: 02
High quality sequence stop: 327.
FEATURES Location/Qualifiers
source 1. .1089
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6484345"

/clone_lib="NIH_MGC_40"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 73 a 59 c 712 g 81 t 164 others
ORIGIN

Query Match 8.1%; Score 105.8; DB 14; Length 1089;
Best Local Similarity 38.6%; Pred. No. 4e-09;
Matches 281; Conservative 0; Mismatches 447; Indels 0; Gaps 0;

```
Qy        575 GGTCGGGTGAGTGCACGCTCGCGCTGGCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCA 634
              ||| | | |                 |                 |
Db        312 GGTAGAGAGGNNNNNNNNNNNNNNNNNNNNNGNNNNNNGNNNNNNNNNNNNNNNNNN 371

Qy        635 TGTTCATCGAGTTCTCCCGGCAGCGCGGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACT 694
                                  |                 | | | |  | |
Db        372 NNNNNNNNNNNNNNNNNNNNNGNNNNNGNNNNNNNNNNGNGNNGNNGNNNNGGGNGGN 431

Qy        695 CGGCTGCAGCCGACGGCACCGGCTGGGGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGT 754
               ||    | | ||    || |||| | ||| |  ||| |  | | || ||    ||
Db        432 NGGGGNNGGGNGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGG 491

Qy        755 TGTCGGATGCGGTGCGGCTGGGGCATCGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCA 814
               |  ||  | || | ||  ||||    ||| |  || || ||   | ||  | | ||
Db        492 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 551

Qy        815 ACCAGGACGGTGCGTCGAATGGGCTGACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGA 874
                ||  || | |  |  | |  | || |  ||   ||   || | ||| |
Db        552 GGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGG 611

Qy        875 TCCGGCAGGCGTTGGCGAACGCGGGGTTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGC 934
                || || |  || |  | |||| |  | || ||  | || |  | || || ||||
Db        612 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 671

Qy        935 ACGGGACGGGCACGACGCTGGGTGATCCGATCGAGGCACAGGCGTTGCTCGCCACGTACG 994
               |||  |||   |  |  |||           | ||    || | |         |  |
Db        672 GGGGGGGGGGGGGGGGGGGGGGNNNNGNGNNGGGGGGGGGGGGGGGNNNNGNNNGGGGG 731

Qy        995 GGCAGCGGGCCGGTGACAGGCCGCTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACA 1054
              ||  | |||  || |   ||  |  | ||| ||||      |  |          |||
Db        732 GGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNN 791

Qy       1055 CCATGGCTGCCGCGGGTGTGGGTGGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGG 1114
                || |  |      | ||||      |  | ||  |  | |  ||  | |||||
Db        792 NGNNGGNGGGNGGNNNGNGNNGGGGGGGGGGGGGGNNGGGNGGGGGGGNGGGGGGGG 851

Qy       1115 TGTTGCCGCGGACGTTGCATGTGGATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGG 1174
                |  |  | ||  |  | ||      |  |  | ||  |  ||  | |||||
Db        852 GGGGGNNGGGGNNGNNGGGGGGGGGGNGGGGNGGGGGGGGGGNGGGGGGGGGGGGGGG 911

Qy       1175 CGGTGCGGCTGCTGACGGAGGCGGTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGC 1234
               || | ||  |  |  || || || |  | ||  |||||  | ||  |||  |   | ||
Db        912 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGNGGGGGGNGGGGGG 971

Qy       1235 GGGCGGGAGTGTCGTCGTTCGGGATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGG 1294
               ||| ||| | |  |  |  |||   || ||   |  | |  | |  |    || || ||
Db        972 GGGGGGGGGGGGNGGGGGGNGGGGNGGGGGGGGGGGGGGGGGGGGNGGGGGGGGNGGGGG 1031

Qy       1295 CGCCGGCG 1302
               |  ||| |
Db       1032 GGGGGGGG 1039
```

RESULT 15
BI416470

LOCUS BI416470 1197 bp mRNA linear EST 15-AUG-2001
DEFINITION hasp001xa01f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xa01f, mRNA sequence.
ACCESSION BI416470
VERSION BI416470.1 GI:15187493
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 1197)
AUTHORS Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026,S-750 07, Uppsala, Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES Location/Qualifiers
source 1. .1197
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="hasp001xa01f"
/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
/note="Vector: pT-Adv; Site_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidon annosum (FP5)."
BASE COUNT 20 a 28 c 1036 g 28 t 85 others
ORIGIN

Query Match 8.1%; Score 105; DB 13; Length 1197;
Best Local Similarity 45.2%; Pred. No. 5.4e-09;
Matches 291; Conservative 0; Mismatches 353; Indels 0; Gaps 0;

```
Qy         653 GGCAGCGCGGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCA 712
               ||  | | |||  |  || ||  ||| || |  ||||     ||  |  |  |  ||
Db         355 GGGGGGGGGGGNGGGGGGGGGGGGGGCGGGGGGGGGGCGGGGGGGGGGGNGNNGGGGGN 414

Qy         713 CCGGCTGGGGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGC 772
                ||  |||| | ||| |  |||       |  || || | ||  |  || | || | ||
Db         415 GGGGGGGGGGGGGGGGGGGGGGGGGNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 474

Qy         773 TGGGGCATCGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGA 832
                ||||    ||| |  || || ||   | ||  | | ||        ||   ||| | |
Db         475 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGTGTGGGGGGG 534

Qy         833 ATGGGCTGACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGA 892
                ||      | |  |    |  ||    || |  || |    ||  ||    || |
Db         535 GNNGGGGNGGNGGGGGGTGGTGGTGGGTGGTGGGGGNGGNGGGGGGGGGGGGNGGGGGG 594

Qy         893 ACGCGGGGTTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGC 952
                | ||||  |   | || ||  | ||  | || || ||||    |||  |||   | |
Db         595 GGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 654

Qy         953 TGGGTGATCCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACA 1012
                ||| |    |  | ||   || | |  |    |   |||   |||  ||
Db         655 GGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGNNGGGGGGGGGGNNGGGGGNGGNNNGG 714

Qy        1013 GGCCGCTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTG 1072
               ||  | | ||  ||||     |          |||      ||  |  | ||| |
Db         715 GGGGGGGGGGGNNGGGGGGGGNGNNNGGGGGGGGGGGGTGGGGGGGGGGGGGGGGGGGG 774

Qy        1073 TGGGTGGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGC 1132
```

```
Db        775 GGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNGGGGGGGGGGGGN 834

Qy       1133 ATGTGGATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGG 1192

Db        835 GGGGGGNGGGGNGGGGGGGGGGNGGGGGGGGGGGNGGGGGGGGNGGGGGGGGGGGGGGGG 894

Qy       1193 AGGCGGTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGT 1252

Db        895 GGGGGGGGGGGGGGGGGGGGGGGTGGGGGGGGGGGGGTGNGGGGGGGGGGGGGGGGGGGGG 954

Qy       1253 TCGGGATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCG 1296

Db        955 GGGGGGGGGGGGNGGNGGGGGGGGGGGGGGGGGGGGGGGGGGG 998
```

Search completed: June 18, 2003, 01:00:11
Job time : 1674.13 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 1786.55 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_3148_4068
Perfect score: 921
Sequence: 1 gtcttcgtgtttcccggaca.....acaacgacacccgccgcttc 921

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		921	100.0	12381	6	AX006889	AX006889 Sequence
2		921	100.0	30690	6	E38020	E38020 Avermectin
3		921	100.0	64957	1	AB032367	AB032367 Streptomy
4		919.4	99.8	31422	6	E38021	E38021 Avermectin
5	c	919.4	99.8	64957	1	AB032367	AB032367 Streptomy
6		677.2	73.5	11096	1	AF275943	AF275943 Streptomy
7		402.4	43.7	65140	6	AX211705	AX211705 Sequence
8		402.4	43.7	123580	1	AF263912	AF263912 Streptomy
9		402.4	43.7	125401	6	AX211739	AX211739 Sequence
10	c	401.4	43.6	104326	1	AB070940	AB070940 Streptomy
11		401	43.5	113193	1	AF357202	AF357202 Streptomy
12		400.8	43.5	27541	6	AX211706	AX211706 Sequence
13		385.4	41.8	20235	1	SERERYAB	M63677 S.erythraea
14		385.4	41.8	20235	6	AR049368	AR049368 Sequence
15		385.4	41.8	20235	6	AR095529	AR095529 Sequence
16		385.4	41.8	20444	1	SEERYABS	X62569 S.erythraea
17		381.6	41.4	104326	1	AB070940	AB070940 Streptomy
18	c	375.6	40.8	84985	1	SNA278573	AJ278573 Streptomy
19		373.8	40.6	24568	1	AF220951	AF220951 Streptomy
20		373.8	40.6	50937	6	AR159871	AR159871 Sequence
21		371.2	40.3	11480	1	STMPKS3ORF	L09654 Streptomyce
22		362.8	39.4	41097	1	AF016585	AF016585 Streptomy
23		355	38.5	78210	1	AB070949	AB070949 Streptomy
24		351	38.1	53784	1	AMM223012	AJ223012 Amycolato
25		351	38.1	53789	6	A69720	A69720 Sequence 3
26		351	38.1	90445	1	AF040570	AF040570 Amycolato
27		349.6	38.0	7788	6	AX089456	AX089456 Sequence
28		349.6	38.0	11219	1	SERERYAA	M63676 Saccharopol
29		349.6	38.0	11219	6	AR049367	AR049367 Sequence
30		349.6	38.0	11219	6	AR095528	AR095528 Sequence
31		349.6	38.0	50000	6	AX089416	AX089416 Sequence
32		349.6	38.0	50000	6	AX089420	AX089420 Sequence
33		348	37.8	80161	1	AY007564	AY007564 Saccharop
34		348	37.8	80161	6	AR165018	AR165018 Sequence
35		347.4	37.7	47981	1	AF263245	AF263245 Micromono
36		347.4	37.7	47981	6	AX112026	AX112026 Sequence
37		345.6	37.5	3185	1	AF411573S4	AF411576 Actinomad
38		345	37.5	28958	6	AR044578	AR044578 Sequence
39		345	37.5	28958	6	I47768	I47768 Sequence 6
40		345	37.5	28958	6	I50958	I50958 Sequence 6
41		345	37.5	28958	6	I63356	I63356 Sequence 6
42		345	37.5	28958	6	I70387	I70387 Sequence 6
43		345	37.5	28958	6	I85639	I85639 Sequence 6
44		345	37.5	28958	6	I88045	I88045 Sequence 4
45		345	37.5	28958	6	I90322	I90322 Sequence 6

ALIGNMENTS

RESULT 1
AX006889
LOCUS AX006889 12381 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 1 from Patent WO0001827.
ACCESSION AX006889
VERSION AX006889.1 GI:9994904
KEYWORDS .
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 12381)
AUTHORS Kellenberger,J.L., Leadlay,P.F., Staunton,J., Mcarthur,H.A. and
Stutzman-Engwall,K.J.
TITLE Polyketides, their preparation, and materials for use therein

```
  JOURNAL    Patent: WO 0001827-A 1 13-JAN-2000;
             KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
             STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PFIZER (US); MCARTHUR
             HAMISH ALASTAIR IRVIN (US); STUTZMAN ENGWALL KIM JONELLE (US)
FEATURES             Location/Qualifiers
     source          1. .12381
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     1884 a   4561 c   4005 g   1931 t
ORIGIN

  Query Match           100.0%;  Score 921;  DB 6;  Length 12381;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-101;
  Matches  921;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3965 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 4024

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4025 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 4084

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4085 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 4144

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4145 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 4204

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4205 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 4264

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4265 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 4324

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4325 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 4384

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4385 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 4444

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4445 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 4504

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4505 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 4564

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4565 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 4624

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4625 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 4684

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4685 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 4744

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4745 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 4804

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       4805 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 4864

Qy        901 GACAACGACACCCGCCGCTTC 921
              |||||||||||||||||||||
Db       4865 GACAACGACACCCGCCGCTTC 4885


RESULT 2
E38020
LOCUS       E38020                 30690 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Avermectin aglycon synthase gene.
ACCESSION   E38020
VERSION     E38020.1  GI:18626909
KEYWORDS    JP 2000245457-A/1.
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 30690)
  AUTHORS   Omura,S. and Ikeda,H.
  TITLE     Avermectin aglycon synthase gene
  JOURNAL   Patent: JP 2000245457-A 1 12-SEP-2000;
            THE KITASATO INSTITUTE
COMMENT     OS   Streptomyces avermitilis
            PN   JP 2000245457-A/1
            PD   12-SEP-2000
            PF   24-FEB-1999 JP 1999046961
            PR
            PI   SATOSHI OMURA,HARUO IKEDA
            PC   C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC   (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(11916)
            FT   CDS             (11971). .(30687).
FEATURES             Location/Qualifiers
     source          1. .30690
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     5355 a   12455 c   8617 g   4263 t
ORIGIN

  Query Match           100.0%;  Score 921;  DB 6;  Length 30690;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-101;
  Matches  921;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3148 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 3207

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3208 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 3267

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3268 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 3327

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3328 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 3387

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3388 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 3447

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3448 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 3507

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3508 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 3567
```

```
Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3568 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 3627

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3628 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 3687

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3688 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 3747

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3748 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 3807

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3808 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 3867

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3868 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 3927

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3928 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 3987

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3988 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 4047

Qy        901 GACAACGACACCCGCCGCTTC 921
              |||||||||||||||||||||
Db       4048 GACAACGACACCCGCCGCTTC 4068
```

RESULT 3
AB032367

```
LOCUS       AB032367               64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
            Sciences, Kitasato University, Microbial chemistry; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .64957
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            101. .12019
                     /gene="aveA1"
     CDS             101. .12019
                     /gene="aveA1"
                     /note="contains a loading module, Ave module 1 and module
                     2
                     multifunctional polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="type I polyketide synthase AVES 1"
                     /protein_id="BAA84474.1"
                     /db_xref="GI:5902891"
                     /translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
                     ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
                     AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
                     VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTR
                     MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
                     MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
                     RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
                     DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
                     LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
                     GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
                     MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
                     GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
                     GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
                     RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
                     GDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
                     LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
                     AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAG
                     LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
                     EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
                     VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
                     AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
                     AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
                     ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
                     PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
                     THHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
                     EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
                     WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
                     LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
                     PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
                     LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGAL
                     ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
                     QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
                     TAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
                     DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSA
                     ARAREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
                     FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
                     AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
                     YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
                     GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
                     GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
                     GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
                     NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
                     WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
                     TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
                     SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
                     ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
                     YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
                     VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
                     MTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
                     LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
                     VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
                     NLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
                     TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY
                     VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGG
                     VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
                     RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
                     EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
                     RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
                     LRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPE
                     DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
                     ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
                     VSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
                     AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
                     LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
```

```
                RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
                RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
                TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
                AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
                SLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLGLS"
gene            12071. .30790
                /gene="aveA2"
CDS             12071. .30790
                /gene="aveA2"
                /note="contains Ave module 3, module 4, module 5 and
                module 6
                multifunctional polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="type I polyketide synthase AVES 2"
                /protein_id="BAA84475.1"
                /db_xref="GI:5902892"
                /translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
                CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
                DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGINP
                EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
                ACQALRSGECTMALAGGASVMATPFVFTEFSRQRGLAADGRCKAFSAAADGTGWSEGV
                GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
                PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
                GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
                GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
                KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
                ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
                NDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALHRL
                LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
                TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
                FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
                TQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
                TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
                PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
                NRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLTAVPADEPIAIV
                GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
                FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
                LTYHDYAARFPTAPAGFEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
                HLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSAAADGTGWGE
                GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAD
                LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
                VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
                SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAADDVPGEAAGDDAGTGGEATGPA
                AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
                IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH
                GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
                PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
                QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
                QGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
                TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
                PQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
                SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
                QHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
                PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
                LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
                IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG
                TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
                TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
                LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
                GTWQGNGLATGQVSEHLRRRGMFAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
                LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA
                AVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ
                FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI
                GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
                RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
                SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
                TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
                AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
                LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
                HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
                DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
                ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
                GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAAL
                LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
                ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
                HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
                QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
                TLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
                PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
                RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
                ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
                TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
                THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
                TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
                IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
                KATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
                PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD
                IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
                TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
                PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
                ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
                YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
                AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL

  Query Match          100.0%;  Score 921;  DB 1;  Length 64957;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-101;
  Matches  921;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3248 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 3307

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3308 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 3367

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3368 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 3427

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3428 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 3487

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3488 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 3547

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3548 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 3607

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3608 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 3667

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3668 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 3727

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3728 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 3787

Qy        541 GGCACCGGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3788 GGCACCGGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 3847

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3848 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 3907

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3908 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 3967
```

```
Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3968 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 4027

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4028 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 4087

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4088 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 4147

Qy        901 GACAACGACACCCGCCGCTTC 921
              |||||||||||||||||||||
Db       4148 GACAACGACACCCGCCGCTTC 4168


RESULT 4
E38021
LOCUS       E38021                31422 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Avermectin aglycon synthase gene.
ACCESSION   E38021
VERSION     E38021.1  GI:18626910
KEYWORDS    JP 2000245457-A/2.
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 31422)
  AUTHORS   Omura,S. and Ikeda,H.
  TITLE     Avermectin aglycon synthase gene
  JOURNAL   Patent: JP 2000245457-A 2 12-SEP-2000;
            THE KITASATO INSTITUTE
COMMENT     OS   Streptomyces avermitilis
            PN   JP 2000245457-A/2
            PD   12-SEP-2000
            PF   24-FEB-1999 JP 1999046961
            PR
            PI   SATOSHI OMURA,HARUO IKEDA
            PC   C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC   (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(14643)
            FT   CDS             (14824). .(31419).
FEATURES             Location/Qualifiers
     source          1. .31422
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     4136 a  10237 c  11677 g   5372 t
ORIGIN

  Query Match             99.8%;  Score 919.4;  DB 6;  Length 31422;
  Best Local Similarity   99.9%;  Pred. No. 2.1e-101;
  Matches  920;  Conservative    0;  Mismatches   1;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16630 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 16689

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16690 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 16749

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16750 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 16809

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16810 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 16869

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16870 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 16929

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16930 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 16989

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16990 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 17049

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17050 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 17109

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17110 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 17169

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17170 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 17229

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17230 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 17289

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17290 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 17349

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db      17350 TACTGGTACCGCAACCTGCACCAGCCTGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 17409

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17410 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 17469

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17470 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 17529

Qy        901 GACAACGACACCCGCCGCTTC 921
              |||||||||||||||||||||
Db      17530 GACAACGACACCCGCCGCTTC 17550


RESULT 5
AB032367/c
LOCUS       AB032367              64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
```

Sciences, Kitasato University, Microbial chemistry; 5-9-1
Shirokane, Minato-ku, Tokyo 108-8641, Japan
(E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
Fax:+81-3-3444-6197)

FEATURES Location/Qualifiers

source 1. .64957
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"

gene 101. .12019
/gene="aveA1"

CDS 101. .12019
/gene="aveA1"
/note="contains a loading module, Ave module 1 and module 2
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 1"
/protein_id="BAA84474.1"
/db_xref="GI:5902891"
/translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTR
MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
GDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAG
LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
THHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGAL
ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
TAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSA
ARAREREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
MTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
NLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY
VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGG
VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
LRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPE
DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
VSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
SLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLGLS"

gene 12071. .30790
/gene="aveA2"

CDS 12071. .30790
/gene="aveA2"
/note="contains Ave module 3, module 4, module 5 and module 6
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 2"
/protein_id="BAA84475.1"
/db_xref="GI:5902892"
/translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGINP
EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
ACQALRSGECTMALAGGASVMATPFVFTEFSRQRGLAADGRCKAFSAAADGTGWSEGV
GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
NDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALHRL
LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
TQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
NRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLTAVPADEPIAIV
GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
LTYHDYAARFPTAPAGFEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
HLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSAAADGTGWGE
GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAD
LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAADDVPGEAAGDDAGTGGEATGPA
AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH
GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
QGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
PQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
QHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG
TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
GTWQGNGLATGQVSEHLRRRGMFAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA

```
                     AVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ
                     FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI
                     GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
                     RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
                     SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
                     TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
                     AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
                     LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
                     HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
                     DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
                     ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
                     GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAAL
                     LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
                     ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
                     HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
                     QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
                     TLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
                     PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
                     RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
                     ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
                     TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
                     THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
                     TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
                     IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
                     KATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
                     PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD
                     IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
                     TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
                     PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
                     ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
                     YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
                     AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL

  Query Match             99.8%;  Score 919.4;  DB 1;  Length 64957;
  Best Local Similarity   99.9%;  Pred. No. 1.7e-101;
  Matches  920;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48228 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 48169

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48168 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 48109

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48108 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 48049

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48048 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 47989

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47988 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 47929

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47928 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 47869

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47868 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 47809

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47808 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 47749

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47748 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 47689

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47688 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 47629

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47628 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 47569

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47568 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 47509

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47508 TACTGGTACCGCAACCTGCACCAGCCTGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 47449

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47448 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 47389

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      47388 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 47329

Qy        901 GACAACGACACCCGCCGCTTC 921
              |||||||||||||||||||||
Db      47328 GACAACGACACCCGCCGCTTC 47308
```

```
RESULT 6
AF275943
LOCUS       AF275943               11096 bp    DNA     linear   BCT 02-SEP-2000
DEFINITION  Streptomyces avermitilis avermectin polyketide synthase gene,
            partial cds.
ACCESSION   AF275943
VERSION     AF275943.1  GI:9964075
KEYWORDS    .
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 11096)
  AUTHORS   Hong,Y.-S. and Lee,J.J.
  TITLE     Targeted Gene Disruption of the avermectin O-methyltransferase gene
            and polyketide synthase gene from Streptomyces avermitilis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11096)
  AUTHORS   Hong,Y.-S. and Lee,J.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2000) Anticancer Agent Research Laboratory, Korea
            Research Institute of Bioscience and Biotechnology, P.O. Box 116,
            YuSong-Gu, Taejon 305-600, South Korea
FEATURES             Location/Qualifiers
     source          1. .11096
                     /organism="Streptomyces avermitilis"
                     /strain="ATCC31271"
                     /db_xref="ATCC:31271"
                     /db_xref="taxon:33903"
     CDS             220. .>11096
                     /note="Avr1"
                     /codon_start=1
                     /transl_table=11
                     /product="avermectin polyketide synthase"
                     /protein_id="AAG09812.1"
                     /db_xref="GI:9964076"
                     /translation="MADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAE
                     FAPYVDWSVEQVLRDSPDAPGLDRVDVVQRTLFAVMISLAALWRSQGVEPCAVLGHSL
                     GEIAAAHVSGGLSLADAARVGDAWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDN
                     PARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEER
                     VVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGP
                     KTFVEMSPHPVLTMGLQELAADLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGH
                     GETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHL
                     GFDSVMGVELRNRLSKATGLRLPVTLILDHTTPAAVAARLRTAALGHLDEDTAPVPDS
```

```
                     PSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQR
                     HAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERA
                     GIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRSAYS
                     FGLEGPAITVDTGCSASLVTLHLACQSLRSGECTFALAGGVSDVHPGMFIEFSRQCGL
                     SVDGRCKAVSAAADGTGWGEGIGLLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQAL
                     ANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMA
                     AAGVGGVIKMVMALREGVLPRTLHVDEPSPQGLAAGAVRLLTEAVPWPGDAAGRLRRA
                     GVSSFGIGGTNAHVILEEAPAAGGCVAGGRVLEGAPGLAISVAESVAAPVAVSAPVAE
                     SVPVPVPVPVPVPVPVSARSEAGLRAQAEALRQYEAVQPDVSLADVGAGLACRQAVLEHH
                     VVILAACTSSSRSAARTTARSSSTARPQARPAPTSARRIGALAAGSGSAALTTGHAPG
                     GDRGGVVFVFPWQGGQWAGMGVRLLCLLRVFARRMQACEEALAPWVDWSVVDILRRDA
                     GDAVWEQADVVQPVLFSVMVSLAALWRSYGIEPNEVLGHSKDEIAAAHIYGALSLKDA
                     AKTVALPPQEVEQLIGERGGRLWVAAVNGPRSTAVSGDAEAVVEVLAYCAGTGVRPRI
                     PVDYASHCPHVQPLREELLELLGDISPQPYGVPFFSTVEGTWLDTTTLDAAYWYRNLH
                     QPVRFSHDVQALADQGHRVLLEVSPHPTLVPAIEDTTEDTARRRHCDRQPPPRRERHP
                     LLPQRLRLDPYYRHRQTHHVAPLLDPPRHLPPPLDAPRPAHLPFQHQHYWLESSQPGA
                     GSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVV
                     PALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNIL
                     TNLHHHGITPIPLPLTTHHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPH
                     PHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLAR
                     TTLLEHPTHTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTL
                     TPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLT
                     TQLQQKGIHLTITTCDTSTPRPTHNNSLNTIPPQHPVTTVIHTGGILDDATLTNLTPT
                     QLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAH
                     HRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIADTER
                     PYVVIADIDWSKIEHTSQTSDLVSAARAREPAVQRPTPPAELHKTLAHQTSADQRAAL
                     LELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFD
                     HPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDF
                     WDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGIS
                     PREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQS
                     TDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMAL
                     AGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRN
                     GHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGT
                     TLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLL
                     PRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEA
                     PADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPG
                     LDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGT
                     GEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQN
                     DNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITGAHLAG
                     ILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSL
                     VISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHT
                     PLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHH
                     NLPNTPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHAQPTPHPHPLDLPTYP
                     FQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLA
                     DHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAAD
                     GKGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAV
                     AVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPAL
                     LNGVVQPLSVLLPGGTGFGKGAGFGKGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGG
                     GREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVP
                     VCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQG
                     WLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDT
                     DTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLTRLESSGDVPAQRS
                     GDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLA
                     GVCGVRDLLLVSRRGPDAPGAEGLRRSWPRGAEVRIVACDVGERREVVRLAGGCSCRV
                     SVFVDFSAVASVTVRLPVASDVRKEAAMAYATVEEFTDYLDPDP"
BASE COUNT     1756 a   4059 c   3528 g   1753 t
ORIGIN

  Query Match           73.5%;  Score 677.2;  DB 1;  Length 11096;
  Best Local Similarity   87.9%;  Pred. No. 2.1e-72;
  Matches  810;  Conservative    0;  Mismatches   48;  Indels   64;  Gaps    4;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              ||||||||||||||| | |||||||||||||||||||||||||||||| |||||||| |
Db       3328 GTCTTCGTGTTTCCCTGGCAGGGTGGGCAGTGGGCCGGGATGGGTGTTCGTCTGCTCTGC 3387

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
                |   | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3388 CTCCTCCGGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 3447

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              |||||||||||||||||||||||||||||||||||||||||||||| |||||||||| ||||
Db       3448 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCAGTGTGGGAGCAGGCC 3507

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              ||||| |||||||||||| |||||||||||||||||||||| |||||||| |||||||||
Db       3508 GATGTAGTCCAGCCTGTACTGTTCAGCGTCATGGTGTCTTTAGCTGCTCTTTGGCGTTCC 3567

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              |||||||||||||||| ||| |||||||||||||||| ||| |||||||||||||||| |||
Db       3568 TACGGTATCGAACCCAACGAGGTCCTTGGCCATTCCAAGGACGAGATCGCGGCCGCACAT 3627

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              | | ||||||||||||||||||||||||||||| ||||||||||||
Db       3628 ATATATGGGGCGCTGAGCCTGAAGGACGCAGCGAAGACTGTTG---------------- 3670

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
                                                      |||| ||| ||||||||||
Db       3671 ----------------------------------------CGCTCCCTCCCCAGGAGGTG 3690

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              |||||||||||||||||    | |  |||||||||||||||| ||||||||||||||||
Db       3691 GAGCAGCTCATTGGTGAGC---GTGGCGGGCGGTTGTGGGTGGCTGCGGTCAACGGCCCC 3747

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db       3748 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGTCGAGGTGCTGGCGTACTGTGCC 3807

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              |||||||||||||||  ||   ||||||||||||||||||||||||||||||||||||||
Db       3808 GGCACCGGGGTGCGGCCC---CGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 3864

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db       3865 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTACGGC 3924

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3925 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 3984

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              |||||||||||||||||||||||||||||| ||||||||| ||| ||| ||||||||||||
Db       3985 TACTGGTACCGCAACCTGCACCAGCCGGTACGTTTCAGCCATGACGTACAGGCCCTGGCG 4044

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              ||| | |||||||||||||| | ||||||||||||||||||||||||||||||||||||
Db       4045 GATCAGGGACACCGCGTCTTGCTAGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 4104

Qy        841 GAAGACACCACCGAAGACACCG-CCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGG 899
              |||||||||||||||||||||| |||||||||||||||||| ||||||||||||||||||
Db       4105 GAAGACACCACCGAAGACACCGCCCGAAGACGTCACTGCGATCGGCAGCCTCCGCCGCGG 4164

Qy        900 CGACAACGACACCCGCCGCTTC 921
              ||| |||||||||||  ||||
Db       4165 CGAGAACGACACCCGCTTCTTC 4186


RESULT 7
AX211705
LOCUS       AX211705               65140 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 1 from Patent WO0159126.
ACCESSION   AX211705
VERSION     AX211705.1  GI:15523937
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 65140)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 1 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
```

```
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .65140
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT     8270 a  25171 c  22273 g   9426 t
ORIGIN

  Query Match           43.7%;  Score 402.4;  DB 6;  Length 65140;
  Best Local Similarity   66.2%;  Pred. No. 7.8e-40;
  Matches  612;  Conservative    0;  Mismatches  306;  Indels    6;  Gaps    2;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              || |||||||| ||||| ||||||  ||||||||  ||||||| |   |   || ||||
Db       8674 GTGTTCGTGTTCCCCGGTCAGGGTTCGCAGTGGGTGGGGATGGGGGCCCAACTCCTCGAT 8733

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
                ||  ||||||||||   ||||||     | ||| | || ||| || ||    ||
Db       8734 GAGTCGGCGGTGTTCGCGGAGCGGATTGCCGAGTGTGCGGCGGCACTCGCCGAGTTCACC 8793

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||  |||| ||  | |||||  | |  | ||| |    ||  || ||||||||  |
Db       8794 GACTGGTCGCTGGTCGATGTGCTGCGGGGTGTGGTGGGTGCGCCGTCGTTGGAGCGGGTC 8853

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              |||||||| ||||| | |  ||||   || |||||||| ||||||||  |||||  |||||
Db       8854 GATGTGGTGCAGCCGGCGTCGTTCGCGGTGATGGTGTCGTTGGCTGCGTTGTGGGGTTCC 8913

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
                ||| |    || || |||||  | || |||||  |||||  |||||||| ||||||
Db       8914 CGTGGTGTGTTGCCGGATGCGGTGGTGGGGCATTCGCAGGGTGAGATCGCTGCCGCGGTG 8973

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||||  || ||||||    ||  ||||| |||   |   || |||||||| || | |||
Db       8974 GTGTCGGGTGCGCTGTCGTTGCGGGACGGGGCGCGGGTGGTGGCGCTGCGGAGTCAGGCC 9033

Qy        361 CTGG---CCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAG 417
              | |     ||  ||  ||| |||||||| |||   || ||  |||||  | |   | | |
Db       9034 ATTGGTCGTGCGTTGGCGGGGCGGGGCGGGATGATGTCCGTCGCGCTGTCGGTGGACGTG 9093

Qy        418 GTGGAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGC 477
              | || | ||    | |  ||    ||   | |||||||| ||| ||||||  || ||||||||||
Db       9094 CTCGAACCGCGGTTGGTCGA---GTTCGAGGGGCGGGTGTCGGTGGCCGCCGTCAACGGC 9150

Qy        478 CCCCGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGT 537
              ||  ||||||   ||  ||| | ||  ||  |||||||| ||||||  | |||   ||
Db       9151 CCGCGCTCCGTCGTGGTCGCCGGCGAGCCCGAGGCGCTGGACGCGCTGCACGCCCGGCTG 9210

Qy        538 GCCGGCACCGGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCAT 597
              |||  |   ||    | |||||||| ||||||  |||| |||||  ||||||||||   | |||
Db       9211 ACCGCCGACGACATCCGGGCCCGCCGGATCGCGGTGGACTACGCCTCGCACTCGCACCAG 9270

Qy        598 GTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCC 657
              ||  ||  |||||  |||||  |||||||| ||||||  |||  |     ||||   ||||
Db       9271 GTCGAGGACCTGCACGAGGAACTGCTGGAGGTGCTGGCGGAGCTGGCGCCGCGCACGTCG 9330

Qy        658 GGCGTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCC 717
              |  |||||||||||||| || |||   |||  |||||||||||||  |      ||||||||
Db       9331 GAGGTGCCGTTCTTCTCGACCGTGACCGGCGACTGGCTGGACACCGCGCGGATGGACGCC 9390

Qy        718 GCCTACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTG 777
              | |||||||| ||||||||||| |   | ||| || |||   || || ||   || ||||
Db       9391 GGCTACTGGTTCCGCAACCTGCGCGGACGGGTGCGGTTCGCGGACGCGGTGGCGGACCTG 9450

Qy        778 GCGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
                ||  |  |   ||||| |||||||| |||||| | |||||    ||        ||
Db       9451 CTGGCGGCGGAGTACCGCGCGTTCGTCGAGGTCAGCTCGCACCCGGTGCTGACGATGGCG 9510

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897
              ||   |||   | ||| ||  |||  |     | |  ||||| ||||| ||| |||||
Db       9511 GTCTTGGACCTGATCGAGGAGGCCGGGGTCACGGCCGTCGCGACCGGCACCCTGCGCCGT 9570

Qy        898 GGCGACAACGACACCCGCCGCTTC 921
              | | |    | | |  ||||||||
Db       9571 GACCAGGGTGGCGCGGGCCGCTTC 9594


RESULT 8
AF263912
LOCUS       AF263912              123580 bp    DNA     linear   BCT 24-MAY-2000
DEFINITION  Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF263912
VERSION     AF263912.1  GI:8050835
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., StrLm,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Biosynthesis of the polyene antifungal antibiotic nystatin in
            Streptomyces noursei ATCC 11455: analysis of the gene cluster and
            deduction of the biosynthetic pathway
  JOURNAL   Chem. Biol. 7 (6), 395-403 (2000)
  MEDLINE   20334850
   PUBMED   10873841
REFERENCE   2  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
            N-7489, Norway
FEATURES             Location/Qualifiers
     source          1. .123580
                     /organism="Streptomyces noursei"
                     /strain="ATCC 11455"
                     /db_xref="ATCC:11455"
                     /db_xref="taxon:1971"
     gene            complement(46. .783)
                     /gene="nysF"
     CDS             complement(46. .783)
                     /gene="nysF"
                     /function="presumably post-translationally modifies ACP
                     domains on PKS"
                     /note="putative 4'-phosphopantheteine transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysF"
                     /protein_id="AAF71762.1"
                     /db_xref="GI:8050836"
                     /translation="MIELILPATVATEAAYDDRPRPGDRLLSSEREVIARAVESRQRE
                     FTTVRHLARRALRRLGHPDRAILPNRRGAPQWPPGIVGSMTHCAGYRAAAVSPAELSA
                     AVSIDAEPNGPLPAGVLNAIALPSERPHLVALAAHRPDVHWDRLLFSAKESVFKAWYP
                     LTQRELDFSEAEIVIDPTQGAFTARLLVPGPLLGGRRVTVFPGRWHSTPALLTTAVHL
                     PAPTPRRDREHRTHLTVNSPLPRPTFG"
     gene            complement(867. .2684)
                     /gene="nysG"
     CDS             complement(867. .2684)
                     /gene="nysG"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysG"
                     /protein_id="AAF71763.1"
                     /db_xref="GI:8050837"
                     /translation="MASPDDLEEERTAPRPARRLVGLLRPHRRSVALAVSMGVGGIVL
                     NAFGPLLLGRVTDLIADGVLGGVPGPAPGIDFAAIGRLLLVLLALYVVASLFMLAQGR
                     LVASAVWRTIHELRRDAREKLTRLPLRHFDRQPAGELLSRTTNDIDNLQQTLQQTLAE
```

```
                LITSIFSLLTMLVLMLVISPSLAVVMLLSVPVSALIAARISKRAQPHYAAQWSANGTL
                NAHVEEVCTGHALIKGFDRRAAAEERFDACNDAVYRAAAKAQFASGAMEPVMMFVANL
                GYVLVAVIGAWKVINGTLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGIASAQRVF
                TLLDAPEQAPDPLRPGTPARAEGRVEFTDVSFRYSPDTPLIENLSLTVEPGSTVAIVG
                PTGAGKTTLGNLLMRFYEPDSGRILLDGTDTATMTRDDLRSRFGLVLQDTWLFGGTIA
                ENIAYGAPGACRADIEEAARATCADRFIRTLPQGYDTVLDDESGTVSAGEKQLLTVAR
                AFLARPAVLVLDEATSSVDTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIVVM
                DAGRIVEQGTHDQLLCAQGLYARLHAARTHTPTAGAAAG"
gene            complement(2662. .4416)
                /gene="nysH"
CDS             complement(2662. .4416)
                /gene="nysH"
                /function="presumably involved in efflux of nystatin"
                /note="putative transporter (ABC family)"
                /codon_start=1
                /transl_table=11
                /product="NysH"
                /protein_id="AAF71764.1"
                /db_xref="GI:8050838"
                /translation="MLLRLLRAQLRPYAWATAALVALQLVQILGTLLLPTLGAALIDQ
                GVVRGDGGRITELGVVMGVVALVQIAAALGAAALAARTATAMGRDLRSALFRRILDFS
                AREIGRFGTPSLLTRSVNDVQQVQNLAQTGFGIVVCAPLMCLGSVLLALRQDVPLALL
                LVALVLVVAVCFGLLLARMGTLYARMQLTLDRLGRLLREAITGVRVVRSFVRDDHERA
                RFAQTNDAFLVVSRRVGRLIATMLPVVLLLMNGFTVALLWTGSHRIDAGRMPIGSLSA
                LLSYLSLILMSVVMLAFVFLSVPRARVCAGRIAEVLDTGSSVAPPAAPQPVRGPAGRI
                ELCAAGYRYPGAEEPVLRDVDLTVEPGERIAVLGSTGSGKTTLLNLVLRLADATEGAV
                RVGGTDVRELTAATLAAAVGFVPQRPYLFSGTVASNLRFGRPDATDEELWEALRVAQA
                ADFVARMPDGLDAEITQGGGNVSGGQRQRLSLARALLRRPEIYLFDDCFSALDQATDA
                ALRTALVPYTAGATVITVAQRISAGRDADRIVVLDRGRVVAQGTHDVLLRTSPTYREI
                ALSQLTEEEAAHGLAGRP"
gene            4714. .5748
                /gene="nysDIII"
CDS             4714. .5748
                /gene="nysDIII"
                /function="presumably involved in mycosamine biosynthesis"
                /note="putative dGDP-mannose-4,6-dehydratase"
                /codon_start=1
                /transl_table=11
                /product="NysDIII"
                /protein_id="AAF71765.1"
                /db_xref="GI:8050839"
                /translation="MSKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKSRV
                SRLASELDFIDGDLMDQGSLVSAVDTVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                VLRMLEAIRMVSGLSTSRTVSPRGQIRFYQASSSEMFGKAAETPQRETTLFHPRSPYG
                AAKAYGHYITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKQGLQDKL
                ALGNLDAVRDWGYAGDYVRAMHLMLQQDAGDDYVIGTGQMHSVRDAVRIAFEHVGLNW
                EDYVVIDPDLVRPAEVEVLCADSAKAQDRLGWKPDVDFPTLMRMMVDSDLAQVSRENQ
                YGDVLLAANW"
gene            5930. .34363
                /gene="nysI"
CDS             5930. .34363
                /gene="nysI"
                /function="responsible for condensation steps 9 to 14 in
                the nystatin polyketide backbone synthesis"
                /note="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="NysI"
                /protein_id="AAF71766.1"
                /db_xref="GI:8050840"
                /translation="MDNEQKLRDYLKLATADLRRTRRRVHKLESAAQEPVAIIGMTCR
                YPGGVRSPEDLWRMVEAGEHGVTPFPTDRGWDLEALAAAPTASGGFLHDAPDFDADFF
                GISPREAVAMDPQQRVVLESAWEAFERAGIDPTSVKGSRTGVFIGAMAQDYRVGPADG
                AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                AGGVTIMSGPGTFIEMGRQGGLSADGRCRSFGDTADGTGWAEGVGILVLERLSDAVRN
                GHEILAVVRGTAVNQDGASNGLTAPNGPSQQQVIQQALVNARLAAGDIDVVEAHGTGT
                TLGDPVEAQALLATYGQNRPADRPLLLGSVKSNLSHTQAAAGVAGVIKMVMAMRHGTL
                PRTLHAEEPTHHVDWSQGAVRLLTDTTDWPATGAPRRAAVSSFGISGTNAHTIIEQAP
                EPQPEDAATAQDDAAGSTPATAPVVPGVVPVLLSGRTPDALRGQAAALRAALDTGRRP
                DLLDLAHSLATTRAGFEHRAVLLATDHPALTDGLTALADADDPAAAPAWITGTTRAET
                RLAVLFTGQGAQRLGAGRELAARFPAFATALDAALDAFTPHLDRPLREVLWGTDAALL
                DRTAYAQPALFAVEVALYRLIESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAATLVA
                ARGRLMQALPDGGAMIAVQASEADVAPLLAGHEDQVAIAAVNGPSAVVLSGAEATVTA
                LAEQLAADGRKTRRLRVSHAFHSPLMEPMLDAFRAVVEDLTLQPPLLPVVSNLTGKPA
                TVAQLTSADYWVDHVRHAVRFADGIDWLARHDTTAFLELGPDGVLSAMAQDCLDAADA
                DAVTLPALRAGRPEEHTLTTALAGLHVHGATLDWTGCFAGTGARRTDLPTYAFQRRRY
                WPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRG
                TTLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRT
                VNVHARPDTGDDTPWTAHATGVLTTADASRQLPASSEQGGTPLAGDPHPALDAAQWPP
                AGAEPLPLDGHYDRLADGGFGYGPVFQGLRAAWRGGDVVYAEVELPEAGRSDAEAFGL
                HPALLDAALHAAPFTGLGERGRGGLPFSWEGVSLHAGGATTLRVRLTPVADDALALTV
                ADGTGAPVLSVDSLVLRSVATQQLDTAAAVARDALFRLDWTPVQPTATDPGPVALLGA
                DPFGLLTHAGFADAPAYPDLAALAAADGPVPTTVVLSLAGTGDDAADPARSAHRCAAE
                ALAAVQTWLDHHERFAAARLVFVTRGATVGRDVAAAAVWGLVRSAQSENPGCFALVDL
                DPDGAVGAAALVAALVSGEPQLAVRGDVLRVARLVRRPLTEVGAGADGTGDGVGDGSG
                VSFSGEGAVLVTGGTGGLGAVLARHLVAEYGVRDLLLVSRSGERAVGAGELVAELAGV
                GARVRVVACDVTDRAAVVELVGGHAVSAVVHAAGVLDDGMVGALTGERLSAVLRPKVD
                AVWHLHEATRGLDLDAFVVFSSLAGVFGSPGQANYAAANAFLDALMTRRRAEGLPGLS
                LAWGPWEQSGGMTGTLTDVDAERLARSGVPPLSVAQGLALFDAAVAGTDATCVPVRLD
                LPVLRARGEVPPLLRSLIRVRARRAAVAGSATAGNLAQRLRRLDEDGRDEMVLDLVRG
                QVALVLGHATGGDVDAGRAFRDLGFDSLTAVELRNRLNTVTGLRLPATLVFDYPTVRH
                LATYVLDELLGTDAEVATVQPAAVAVADDPIVIVGMACRYPGGVSSPEDLWRVLTEGT
                DAVSGFPTNRGWDVESLYHPDPHDPGTSYTRSGGFLHEAGEFDPGFFGMSPREALATD
                SQQRLLLESSWEAIERAGIDPVSLRGSRTGVFAGVMYSDYSAMLASPEFEGFQGSGSS
                PSLASGRVAYTLGLEGPAVTVDTACSSSLVAMHWAMQALRSGECGLALAGGVTVMSTP
                AVFVDFARQRGLSPDGRCKAFADAADGVGWSEGVGVLVLERQSDAVRNGHEILAVVRG
                SAVNQDGASNGLTAPNGPSQQRVIRQALASGGLTAGDVDVVEAHGTGTTLGDPIEAQA
                LLATYGRDREPERPLLLGSVKSNLGHTQAAAGVAGVIKMVLAMRHGVVPRTLHVDAPS
                SHVDWSEGAVELLSEQAAWPETGRVRRAGVSSFGISGTNVHVIVEQAPGAKAIAAAGA
                ARRTPGAVPVLLSGRGRSALRGQAARLLGHLQARPDAELVDVALSLATTRSRFEQRAA
                VVAQDRDQLIASLGALAADRPDPAVVEGEAAGRGRTAVLFTGQGSQRAAMGRELHEVQ
                PEFAAAFDAVCAVFDPLLDRPLREVVFAEDGSDEAALLDETGWTQPALFAVEVALFRL
                VESWGVRPDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARATLMQALPTGGAMIAIQA
                TEDEIAAHLDDTVAIAAVNGPQSVVISGDEEAAETIAATFAERGRKTKRLRVSHAFHS
                PRMDGMLDAFRIVAEGLTYRAPRIPLVSDLTGRRADDAEVCTAEYWVRHVREAVRFAD
                CVRTLRDAGATTFLELGSDGLLTAMAEDTLGDDHDAELVPMLRAGRAEELAAATALAR
                LQVRGVDVDWAAYLAGTGARRTDLPTYAFQHAYYWPQLPTPAAALAAADPADQQLWAA
                VERGDARELADILGLGEQDLTPLDSLLPALTSWRRGNQEKHLLDTLRYRVEWTRLSKP
                TAPVLDGTWLLVASDATAADQPALLDGLADALGSHGARVRRLLLDDSCADRAVLAERL
                ARTADVDAATQVLSVLPLDERDADDCPPLTRGLALTVALVQALADTGAQGRLWTATRG
                AVSTNPADPVTHPVQAAAWGLGRGVALEHPRLWGGLVDLPQVFDERAGQRLAGILAVK
                DAPDGEDQVALRATGVSGRRLVRHTVEALPTAAEFTATGTVLITGGTGGLGAEVARWL
                ARAGAQHLVLTSRRGPDAPGAAELRAELEGYGPSVSVVACDVADRDALAAVLTALPEE
                LPLTGVVHTAGVGHYGPLDTLSTAEFAGLTAAKLAGAAHLDALLADRELDFFVLFGSI
                AGVWGSGNQSAYGAANAYLDALALHRRARGLAATSVAWGPWAEAGMAADDAVSETLRR
                QGLGLLDPAPAMTELRRAVVRQDVTVTVADVDWQRYAPLFTSARPSALIAGLPEVRAL
                AADERTEQDATGASEVVTRVRALAEPEQLRLLTDLVRTESATVLGHSSADAVEPEGRAF
                RDVGFDSLTAVELRKRLGAATGLSLPSTMVFDYPTPLELAQYLRAEILGAVLEVAGPV
                ATGGADDEPIAIIGMACRFPGGVSSPEQLWDLVASGTDAISEFPVNRGWQTGHLFDPD
                PDRPGTTYSTQGGFLHEADEFDPTFFGISPREALVMDPQQRLLLETTWESFERAGIRP
                ETLRSTLTGTFVGSSYQEYGLGAGDGTEGHMVTGSSPSVLSGRLSYVFGLEGPAVTVD
                TACSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNPFIAFSRQRALAKDGRCKAFS
                DDADGMTLAEGVGVVLVERLSDAQRNGHPVLAVLRGSAINQDGASNGLTAPNGPSQQR
                VIRQALANARLAPGDIDALEAHGTGTPLGDPIEAQALFATYGRDRDPESALLLGSVKS
                NIGHTQSAAGIASVIKMVMALRHSELPPTLHADAPSSHVDWSAGTVRLLTQARAWPET
                GRPRRAAVSSFGISGTNAHVLLEQAPVADTPAEERPAVAPVPIAAGVVPWVVTARSAA
                ALRGQAERLLAHAETVGTALPAAGPLDIGLSLVSARARFEHRAVVVPPAGTDPLAALR
                AVATDGPSPVVARGVADVEGRTVFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAAL
                AEFTDWSLVDVLRGVVGAPSLERVDVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQG
                EIAAAVVSGALSLRDGARVVALRSQAIGRALAGRGGMMSVALSVDVLEPRLVEFEGRV
                SVAAVNGPRSVVVAGEPEALDALHARLTADDIRARRIAVDYASHSHQVEDLHEELLEV
                LAELAPRTSEVPFFSTVTGDWLDTARMDAGYWFRNLRGRVRFADAVADLLAAEYRAFV
                EVSSHPVLSMAVQEAIDEAGVPAVAAGTLRRDQGGTDRFLLSAAEVFVRGVDVDWAGL
                FEGTGASRIDLPTYAFQHEHLWAVPPAPEAVAAADPDDAAFWTAVEDGDVSALTAALG
                TDEDSVAAVLPALTSWRRARRDRSTVDAWRYRVAWKPLGGTLPHPSLTGTWLLVTADG
                IDDTDVAGALETYGAEVRRLVLDEECVDRAVLRERLAGAEDVTGIVSVLAAAERTDAV
                PGTSLVLGTALTVALIQALGDAEIDAPVWALTRGAVSTGRADELTAPVQAQVTGIGWT
                AALEHPQRWGGTLDLPAALDARAAQRLAAVLSGALGSDDQLAIRPSGVFTRRIVRAEA
```

```
Query Match          43.7%;  Score 402.4;  DB 1;  Length 123580;
Best Local Similarity  66.2%;  Pred. No. 6.6e-40;
Matches  612;  Conservative    0;  Mismatches  306;  Indels    6;  Gaps    2;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              || |||||||| ||||| ||||||  ||||||||  |||||||| |  |  || ||||
Db      68527 GTGTTCGTGTTCCCCGGTCAGGGTTCGCAGTGGGTGGGGATGGGGGCCCAACTCCTCGAT 68586
```

```
Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
Db      68587 GAGTCGGCGGTGTTCGCGGAGCGGATTGCCGAGTGTGCGGCGGCACTCGCCGAGTTCACC 68646

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
Db      68647 GACTGGTCGCTGGTCGATGTGCTGCGGGGTGTGGTGGGTGCGCCGTCGTTGGAGCGGGTC 68706

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
Db      68707 GATGTGGTGCAGCCGGCGTCGTTCGCGGTGATGGTGTCGTTGGCTGCGTTGTGGGGTTCC 68766

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
Db      68767 CGTGGTGTGTTGCCGGATGCGGTGGTGGGGCATTCGCAGGGTGAGATCGCTGCCGCGGTG 68826

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
Db      68827 GTGTCGGGTGCGCTGTCGTTGCGGGACGGGGCGCGGGTGGTGGCGCTGCGGAGTCAGGCC 68886

Qy        361 CTGG---CCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAG 417
Db      68887 ATTGGTCGTGCGTTGGCGGGGCGGGGCGGGATGATGTCCGTCGCGCTGTCGGTGGACGTG 68946

Qy        418 GTGGAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGC 477
Db      68947 CTCGAACCGCGGTTGGTCGA---GTTCGAGGGGCGGGTGTCGGTGGCCGCCGTCAACGGC 69003

Qy        478 CCCCGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGT 537
Db      69004 CCGCGCTCCGTCGTGGTCGCCGGCGAGCCCGAGGCGCTGGACGCGCTGCACGCCCGGCTG 69063

Qy        538 GCCGGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCAT 597
Db      69064 ACCGCCGACGACATCCGGGCCCGCCGGATCGCGGTGGACTACGCCTCGCACTCGCACCAG 69123

Qy        598 GTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCC 657
Db      69124 GTCGAGGACCTGCACGAGGAACTGCTGGAGGTGCTGGCGGAGCTGGCGCCGCGCACGTCG 69183

Qy        658 GGCGTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCC 717
Db      69184 GAGGTGCCGTTCTTCTCGACCGTGACCGGCGACTGGCTGGACACCGCGCGGATGGACGCC 69243

Qy        718 GCCTACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTG 777
Db      69244 GGCTACTGGTTCCGCAACCTGCGCGGACGGGTGCGGTTCGCGGACGCGGTGGCGGACCTG 69303

Qy        778 GCGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
Db      69304 CTGGCGGCGGAGTACCGCGCGTTCGTCGAGGTCAGCTCGCACCCGGTGCTGACGATGGCG 69363

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897
Db      69364 GTCTTGGACCTGATCGAGGAGGCCGGGGTCACGGCCGTCGCGACCGGCACCCTGCGCCGT 69423

Qy        898 GGCGACAACGACACCCGCCGCTTC 921
Db      69424 GACCAGGGTGGCGCGGGCCGCTTC 69447
```

```
RESULT 9
AX211739
LOCUS       AX211739               125401 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 35 from Patent WO0159126.
ACCESSION   AX211739
VERSION     AX211739.1  GI:15523950
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 125401)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 35 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .125401
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT    15664 a  49692 c  42871 g  17174 t
ORIGIN
```

Query Match 43.7%; Score 402.4; DB 6; Length 125401;
Best Local Similarity 66.2%; Pred. No. 6.6e-40;
Matches 612; Conservative 0; Mismatches 306; Indels 6; Gaps 2;

```
Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
Db      68935 GTGTTCGTGTTCCCCGGTCAGGGTTCGCAGTGGGTGGGGATGGGGGCCCAACTCCTCGAT 68994

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
Db      68995 GAGTCGGCGGTGTTCGCGGAGCGGATTGCCGAGTGTGCGGCGGCACTCGCCGAGTTCACC 69054

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
Db      69055 GACTGGTCGCTGGTCGATGTGCTGCGGGGTGTGGTGGGTGCGCCGTCGTTGGAGCGGGTC 69114

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
Db      69115 GATGTGGTGCAGCCGGCGTCGTTCGCGGTGATGGTGTCGTTGGCTGCGTTGTGGGGTTCC 69174

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
Db      69175 CGTGGTGTGTTGCCGGATGCGGTGGTGGGGCATTCGCAGGGTGAGATCGCTGCCGCGGTG 69234

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
Db      69235 GTGTCGGGTGCGCTGTCGTTGCGGGACGGGGCGCGGGTGGTGGCGCTGCGGAGTCAGGCC 69294

Qy        361 CTGG---CCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAG 417
Db      69295 ATTGGTCGTGCGTTGGCGGGGCGGGGCGGGATGATGTCCGTCGCGCTGTCGGTGGACGTG 69354

Qy        418 GTGGAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGC 477
Db      69355 CTCGAACCGCGGTTGGTCGA---GTTCGAGGGGCGGGTGTCGGTGGCCGCCGTCAACGGC 69411

Qy        478 CCCCGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGT 537
Db      69412 CCGCGCTCCGTCGTGGTCGCCGGCGAGCCCGAGGCGCTGGACGCGCTGCACGCCCGGCTG 69471

Qy        538 GCCGGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCAT 597
Db      69472 ACCGCCGACGACATCCGGGCCCGCCGGATCGCGGTGGACTACGCCTCGCACTCGCACCAG 69531

Qy        598 GTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCC 657
Db      69532 GTCGAGGACCTGCACGAGGAACTGCTGGAGGTGCTGGCGGAGCTGGCGCCGCGCACGTCG 69591

Qy        658 GGCGTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCC 717
Db      69592 GAGGTGCCGTTCTTCTCGACCGTGACCGGCGACTGGCTGGACACCGCGCGGATGGACGCC 69651
```

```
Qy        718 GCCTACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTG 777
              |||||||| |||||||||||| |   | ||| || |||   || || ||   || ||||
Db      69652 GGCTACTGGTTCCGCAACCTGCGCGGACGGGTGCGGTTCGCGGACGCGGTGGCGGACCTG 69711

Qy        778 GCGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
              ||  |  |  |||||  |||||||| |||||| | |||||    ||        ||
Db      69712 CTGGCGGCGGAGTACCGCGCGTTCGTCGAGGTCAGCTCGCACCCGGTGCTGACGATGGCG 69771

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897
              ||   |||   | ||| ||  |||  |    | |  ||||| ||||| ||| |||||
Db      69772 GTCTTGGACCTGATCGAGGAGGCCGGGGTCACGGCCGTCGCGACCGGCACCCTGCGCCGT 69831

Qy        898 GGCGACAACGACACCCGCCGCTTC 921
              | | |    | | |  ||||||||
Db      69832 GACCAGGGTGGCGCGGGCCGCTTC 69855


RESULT 10
AB070940/c
LOCUS       AB070940              104326 bp    DNA     linear   BCT 22-JAN-2002
DEFINITION  Streptomyces avermitilis oligomycin biosynthetic gene cluster.
ACCESSION   AB070940
VERSION     AB070940.1  GI:15823967
KEYWORDS    .
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
  TITLE     Genome sequence of an industrial microorganism Streptomyces
            avermitilis: deducing the ability of producing secondary
            metabolites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
  MEDLINE   21477403
REFERENCE   2  (bases 1 to 104326)
  AUTHORS   Ikeda,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .104326
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            complement(1. .1134)
                     /gene="mcmA"
     CDS             complement(1. .1134)
                     /gene="mcmA"
                     /note="methylmalonyl-CoA mutase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB69185.1"
                     /db_xref="GI:15823968"
                     /translation="MTGYSTAADANARFRQLVGSGTRAVPVVFDLPTRAGLDSDAPAA
                     RGLVGRRGVAVDSIDDMRVLFGGLRLDRVVPSLAVDAPAAPLLVLYQLVAEEHGAPAQ
                     RLAGAVRGDVFKEFLARGPGVFPLRPSLRLAMNVFAYCQAELPRWQCLSVCGHHLAGT
                     GADPALEVAFTVAGGIEYLRAAVTAGLDVRTVAPRLTLCLVGGTTGRENRAKLRATRR
                     VWTRTANERFGSGRTAPRLYVRTPRDGLATLACTEDIEAGATALVDRIERLGGTLTVL
                     EHGLAGQLLGRPHGDRPATPRPHPPHQDSLDLQARQTERLAKLRAWRVQPAVDEALLR
                     LRKAAQGDDNVLYPMREALAARATVGEVCDALRSTWGTGPSAT"
     gene            complement(1131. .1886)
                     /gene="pptA"
     CDS             complement(1131. .1886)
                     /gene="pptA"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphopantheteinyl transferase"
                     /protein_id="BAB69186.1"
                     /db_xref="GI:15823969"
                     /translation="MRAHEVFADVPVSSLHPREAALVQDAVESRRREFATGRHCARRA
                     LASLGVPPGPLLATPRGAPRWPAGVCGAITHCAGYRAAAVARAPGIRAVGIDAEPDRP
                     LPEGVLEAVALPAEQFQVKEHTARHPGVVWGKLLFSAKESVYKAWFPATGRPLGFEEA
                     LIDLDPVSSTFRVRVLPTALRAAEPELRDLHGRWLVRHGIIVTAVVVDGPGRDTPHAP
                     SPYDRTKHSAPTGRSADRPGGRPTDRATDRPTA"
gene                 complement(1910. .2875)
                     /gene="simX4 homolog"
CDS                  complement(1910. .2875)
                     /gene="simX4 homolog"
                     /codon_start=1
                     /transl_table=11
                     /product="SimX4 homolog"
                     /protein_id="BAB69187.1"
                     /db_xref="GI:15823970"
                     /translation="MSQDRHDGSTAALGGPSHGGRLLAVSDIHVRHRQNRDIVQDLHP
                     GCAEDWLIVAGDVAESVADITWALSLLRDRFAKVIWAPGNHELWTAPGDPDELRGERK
                     YRHLVDLCRALGVVTPEDPYPVWHGAGGPVAVAPLFLLYDYSFLPPGAWSKSHALRLA
                     DDAGVICTDEFLLHPDPYPSREAWCRARVRATRARLDALPDDLPTVLVNHFPLVREPT
                     RVLRHPEFALWCGTTATADWHLRYRAAAVVHGHLHIPRTIRCDDVPHHEVSLGYPREW
                     QRYGRRPDWPVQILTPAPAPAPAPVPVPVPAPPGHRTASPADRRP"
gene                 3213. .3971
                     /gene="olmC"
CDS                  3213. .3971
                     /gene="olmC"
                     /codon_start=1
                     /transl_table=11
                     /product="thioesterase"
                     /protein_id="BAB69188.1"
                     /db_xref="GI:15823971"
                     /translation="MSANRVDGNLWVRQFHPAPTAATKLVCLPHAGGSASYFLPVSKA
                     LAPGVDVLAIQYPGRQDRRHEKCIDNIAELADAVTAVLEPLADRPLVLFGHSMGATLA
                     FEVALRLERKGIVPLALFASGRRAPSRHRDESVHLRGDDEIIAELKTLSGTRSEVLGD
                     EEILRMVLPAIRSDYKAAETYRYAGGARLTAPIHAHFGVEDPRVSLDEAQAWADHTTG
                     HFELRSHPGGHFYLNDQAPRVIEEIARVSSSVAS"
gene                 4643. .7438
                     /gene="olmRI"
CDS                  4643. .7438
                     /gene="olmRI"
                     /codon_start=1
                     /transl_table=11
                     /product="luxR-family transcriptional regulatory protein"
                     /protein_id="BAB69189.1"
                     /db_xref="GI:15823972"
                     /translation="MLIGRGQEIDRLRAAYSSVRSTGRSRAVLVESAAGCGKSEILED
                     LVRHVESLGLIALRATAIAEERNTPLSVLRSLVNGAPFPDDLVRRFHTMSDAAARMIS
                     TEGRDKARRDAAWYELMRTFCTEVRALAQRAPLVVAVDDIHHGDADSLRQLLHLARHC
                     RNVPVLLTFTQSLAGGPADRTFGAELLRQPNLEHIRLPRFCRQDIADMLAEQGRADAG
                     QGTVERYFAVSGGNPLLLRALLEDARLEPDDTVDAQDVGPRAGQAYGRAVLACLDRGG
                     EKAHEVAAGLAVLGGAFTHELLSRLLDISAAALTENLRALRAAGIVVGTSFQHPAAKA
                     AVLDSLEPSHRLRLHQRSAALLHASGAPAALAAEHLHEAGQADGAWAVGVLQQAAEQA
                     LLADEARRAVSYLELAHAASADPRGRTEIKIRLGSVTRRVSAADAERHVDDALRGFGA
                     GLLTPSDTGGLAALLLAHGRLEQARSVLEQGRRPGDGDADRTAARRQLAEIRDSVLNR
                     GAWCAGTRAAWHENPAGRPSTGRAGTGLRAWLSGSHAALLAQRIAGTEGLLEITHLTD
                     ATFDLVLNAVWTLLALGAVERALHWCDTFLREAKRRDAPGWEAVFAAVRAEGALRGGN
                     LAEAERDARRCFAVLPERHRSALEGSAVSYLVTVQIARGKYEDAARQLSRPMADSLPD
                     TIHWLDYLRARGLYYMATRQYHSALRDFHEVGRLAERWGTDWPALLPWRTDTAEALLR
                     LGEREQGLRLATEQLAMVQDGAPRIKGISLRLQGLASEPAARLGLLNRAVVELSRADD
                     RLELSRTLFDLADAHRETRQTRQAAMVLRRARQLAKSCGAQPPRDRIGSPDAAGASGE
                     LFGEGPERNADWGHLSESERRVAGLAASGYTNREISVRLYITMSTVEQHLTRVYRKLK
                     INGREELLAEARAGVQV"
gene                 7658. .10486
                     /gene="olmRII"
CDS                  7658. .10486
                     /gene="olmRII"
                     /codon_start=1
                     /transl_table=11
                     /product="luxR-family transcriptional regulatory protein"
                     /protein_id="BAB69190.1"
                     /db_xref="GI:15823973"
                     /translation="MLASAGQGEGGVVLVNAGPGVGRTALLDAFLRKATESGARVCAA
                     AGSSAETEAEFGVVGQLFPPTGPLGAGFRLARAIGSRDAGPGLDADRLFDALCQELRR
                     RPTVIAVDDAQLADAASLRFLRYLVRRLRTAPVMMVLTEPTGPDALPPSFQAELLRHP
```

```
                QCRRLGLLPLSRDGVTQVVSAQVGSDRVPRLAADFHAISGGNPLLVHALLADHRAAHR
                LSPVPSVPTVPSVSSASSAASASSAASAASAVSAAPAASAAFAKAALAWAYRDEPVLF
                DVACGFAVLGESATPARVACLVDRGADAVDQVMTALDTAGLLAGGTFRSALVGKALLE
                HVDVETRAELRRRAARLLHGDGAPATVVARHLLAAPVAEPWAGRTLLAAAEKAMQEGR
                EELSLDCLRLAGQSATTEQERAAVVMARVKIGWEIDPRLVSPWLDELVAALRRGDLRS
                GDAAWIVKHLAWQDRFPEAVESLSTLAAAGDAGEPGVEAELHFVRHWLRHSCPPLEAS
                VPGPSNPTRPRTPRLRRVSPASYAGEVLGKVLTEGPDDHTVAMAEEILRGCRYSETVV
                EALEAALLALVYAGRPDRAAFWCDGLLQQAGTRSTGTSVAILSGIRAEIALRQGALPA
                VEQYARSALSAISLPGWGVAVGSPLAALVHSTTASGMSGVAGAWLNQAVPDGMFRTRH
                GLLYTHARGHYHLAVDRPAAALDDFLTCGGLAKEWGMDVPTFLPWRTSAAQAQLALGN
                PGQARALVREQLGRPGGSTARARGVALRLLAATGEPGRRPALLRESVNLLESCGDQVE
                LLRALADRCQALDHVGAPAKARMTARHARNVADACGGDALFRKLFRDEVPDREGDPDG
                LTGEEQGSASLTAAERRVTALAALGHSNREIGRKLFITKSTVEQHLTRVYRKLGVRNR
                ADLGDLLAGTNLGAGTPAMAMASSAVAG"
CDS             10933. .11271
                /codon_start=1
                /transl_table=11
                /product="P450-like protein"
                /protein_id="BAB69191.1"
                /db_xref="GI:15823974"
                /translation="MEVGGQTIRAGEGVLCTLSTANRDEGAFTDPDELDLRRDARSHL
                SFGFGIHQCIGQSLARAEPQIALETLLRRLPGLRLAASFEELRFRDGVGFYGVKELPD
                PCPALPPDRP"
gene            11458. .29898
                /gene="olmA1"
CDS             11458. .29898
                /gene="olmA1"
                /codon_start=1
                /transl_table=11
                /product="modular polyketide synthase"
                /protein_id="BAB69192.1"
                /db_xref="GI:15823975"
                /translation="MACRLPGAAGPAEFLALLRNGEDAVGAAPGDRRYAPRRGGFLDS
                VDEFDAGFFGISPREAAAMDPQQRLVLELSWEALEDAGIPPARLGGSGAGVFVGAIAD
                DYATLSRAAGADAATPQTTTGLNRGIIANRVSYTLGLRGPSFTVDSGQSSSLVAVHLA
                AESLRRGECSVALAGGVNLILAEDSTVAVERFGALSPDGYCYTFDARANGYVRGEGGG
                VVVLKPLADAIADGDDIHCILAGSAVNNDGGGEGLTVPDRAGQEAVLTAAYEQAGISP
                AAVGYVELHGTGTPTGDPVEAAAVGAVLGAGRPAGQPLLVGSVKTNIGHLEGAAGIAG
                LLKAVLAVRHREIPATLHHASPNPAIPMAELGIGVNTALTPWPEQAGPLVVGVSSFGM
                GGTNCHVVLTEPAGYDAAARPATAGIRPAGTPAPLLVSGRDEQALRAQARKLRDHLDT
                HPGLPLRDVAHTLAAARAPLARRAVLLAADREDVSAGLARLADGTPGADLVRGTVSPG
                ALAFVFTGQGSQRPGMTAELYRTSAEYAAALDEVCASLDPWLRRPLRDVLFAAEGSDE
                AALLDRTEFTQPALFAVEVALFRYAEHCGLTPRFLLGHSVGELAAAHVAGVLSLADAC
                TLVAARGRLMQDRPATGAMVSVQATEDELAPFVSESVTLAALNGPASTVIAGDEDAVL
                AVAAHWAARGRKTKRLRVSHAFHSPHMDGMLTEFRRIAAGLTFGAPRIPIVSNETGTL
                LTEADARSPEYWVRHARAAVRFLDGVRLLEDQGVTTLLELGPDGVLSAMARDCLRDAG
                TVCVPLLRGRDRTEPEALLAALAGLYVRGVPMDWERTATEAGARRVPLPTYAFQRRRH
                WLPEPSASASRPVVNGGGHAGQAGHAGQAGRAGQAGHAGHDGEPAPAVRDRTTPELVR
                DRTTLELVRDATAQVLGHASPDDVATHEAFRELGFSSLMLAELGERLTEATGRRVPTT
                LLFDHPTPDALARRLEADGAERTDAGPVSPPAAAAHDDDPIVIVGMGCRLPGGVRSPE
                DLWRFVAAGGDAISALPVDRGWEFAADFPGEGGFLADVAEFDAGFFGISPREALAMDP
                QQRLLLETSWEALERAGVDALKLRGSRTGVFVGASPSEYGPRLHETSDADGHVLTGTA
                PSVLSGRIAYVLGLEGPALTVDTACSSSLVALHLAAQALRGGECDLALAGGVTVMATS
                GMFAEFARQGGLAPDGRCKAFADGADGTGWGEGVGMLAVQRLSDAVRDGRPVLAVVRG
                SAVNSDGASNGLTAPNGPSQQRVIRQALASAGLSAADVDMVEAHGTGTSLGDPIEAQA
                LLATYGQDRPDDRPLWLGSVKSNIGHTQAAAGVAGVIKSVLALRNGLLPKTLHVDEPT
                PEVDWSAGAVELLTEGREWPETDGPRRAGVSAFGISGTNAHVILEQAPSVEEPVTAPA
                DDGLLVPWVVSARSEEALRAQARRLADHVRESGLRPADVGLSLAVTRAGLEHRAVLVA
                EDEAGFLAHLDALAEGRTPAGVVVEGGTAFLFSGQGSQRAGMGRELYSAHPVFAAAFD
                EVCAAFDGLLPGALREVVFGGAGLDRTEWAQPALFALEVALFELVSSWGVRADVLLGH
                SIGELAAAYVAGVWSLADACRIVAARGRLMQALPEGGAMVAVEAAEDELPDLPGGVSV
                AAVNGPRSLVLSGDEEPVTALAQAFAEQGRRTKQLTVSHAFHSARMEPMLAEFAETLA
                AVEFRPPRIPVVSNVTGEMAGDEPTTPAYWVRHVREAVRFADGVGAAHAAGVRRYLEL
                GPDGTLTTLAHDTAHDAGRDVAAVPALRGDQPETRTVVTAVGRLYATGLTVDWPAFFA
                PYGARPTELPTYAFQHRRYWLEPVAPSSAPPANPLRYRTAWASVPDGPRPSLSGTWAV
                VVPASGSYEDIVAGVTAALRDHGARPVVVEVDPAGPERADVADTLRAALGEDAGGVVS
                LLGTAGSLSSGVAATAVLVQAWTDLGGGARLWCLTRGAVSVSPSDPLAAPAQAQLWGM
                GRVAALEHPERWGGLVDLPAVPDDRAWTRLCAVLAGGSGEDQVAVRAAGLFARRLGRA
                GTRADAAPAGPSPWRTDGTVLVTGGTGALGAHLAQWLAAAGAAHVLLTSRRGPDADGA
                AELTARLRETGTEVTVAACDVTDRDALAELLSGLPADRPLTGVVHAAGVLDDGVLDSL
                TPDRFAAVAGPKVTGAWHLHELTRDLDLSAFVLFSSFAGTVGLAGQANYAAANAYLDA
                LAVHRAQLGLPATAVAWGSWSGAGMAADTEAARRQLARTGLVPLAPDAALAALGQVMD
                DGETAVTVADVDWDRFAAGFGPGRPSPLLTGVPEVRRIEERETGRAGAPAETDRLGGL
```

```
 Query Match             43.6%;  Score 401.4;  DB 1;  Length 104326;
 Best Local Similarity   66.2%;  Pred. No. 9.1e-40;
 Matches  611;  Conservative    0;  Mismatches  306;  Indels    6;  Gaps    2;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              || ||||| || ||||| ||||| |  |||||||| |||||||  | |  ||||| | |
Db      93926 GTGTTCGTCTTCCCCGGCCAGGGCGCCCAGTGGGCGGGGATGGCGGCGGAACTGCTGGTC 93867

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              ||| | |||||||||        |||||  ||| ||||||||||||| |||   ||||  ||
Db      93866 TCCGCACCGGTGTTCCGGGCCCGGATCGAGGAGTGCGAGGAGGCGCTGCGCCCGTACGTC 93807

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              |||||||| || | | |||||||  | |||  ||||      ||   ||||| ||||
Db      93806 GACTGGTCGCTGCTCGCCGTCCTGCGGCAGGAGCCGGGCACCCCGCCCTGGGACCGGGTG 93747

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              || ||||| ||||| | |||||     || ||||||||  |||| ||  ||||||||||
Db      93746 GACGTGGTGCAGCCCGCGCTGTGGGCGGTGATGGTGTCGCTGGCGGCGCTGTGGCGTTCG 93687

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              |||||  |||||||  |||||||| || || ||  || |||||||||||||||||||||    |
Db      93686 TACGGCGTCGAACCGGACGCGGTGCTCGGGCACTCGCAGGGCGAGATCGCGGCCGCCTGT 93627

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              || || |||||||||  |||    ||||  ||   |  ||  || ||||||  | |||||
Db      93626 GTCTGCGGGGCGCTGGACCTCGCCGACGGCGCCCGGGTCGTCGCCCTGCGCGCCAGGGCG 93567

Qy        361 CTGGCCGCTGTGC---GGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAG 417
              | | | |   ||    ||| |||||||| ||||  || |||  |||||||  |   || |
Db      93566 ATCGGCTCGACGCTCGCGGGGCGGGGCGGGATGGTGTCGGTGGCGCTGCCGTCGGCGGCG 93507

Qy        418 GTGGAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGC 477
              |    | |||| || | |  ||||| |||||||  | | ||||||  ||||||||||
Db      93506 GCCCGGGAGCTGATCGCCG---GGTGGCCGGGGCGCGTCTCGGTGGCCTCGGTCAACGGT 93450

Qy        478 CCCCGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGT 537
              |||  | |||||  ||  | || ||| |||| ||| | |||||  |||| ||   |||
Db      93449 CCCGTCAGCACCGTGGTGGCCGGTGATCCCGACGCGCTCGACGAACTGCTCGCCCGCTGC 93390

Qy        538 GCCGGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCAT 597
              |||| |  ||| |||||||||||  || ||||||||||||| |||||  |||     ||||
Db      93389 GCCGACGACGGCGTGCGGGCCCGTCGCATCCCGGTCGACTACGCCTCCCACACATCCCAA 93330

Qy        598 GTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCC 657
              |||  |    |  ||  ||  ||  |    |||| ||  |  | |||||   ||||
Db      93329 GTGGCGGAGATCGAGGGCGACCTGGCCGCCGAACTGGCGGGGCTGACGCCGCGTTCGTCG 93270

Qy        658 GGCGTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCC 717
              |   | ||||||||||||||  |   ||  ||  ||| |||||| | |  |||||||||
Db      93269 GCGATCCCGTTCTTCTCCACCCTCACCGGGGCCCCGCTCGACACCGCGAAGCTGGACGCC 93210

Qy        718 GCCTACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTG 777
              | |||||||| |  |||||||  ||  | ||    |||   |  ||   ||  ||
Db      93209 GGCTACTGGTATCAGAACCTGCGTCACACCGTGGAGTTCGAGGCGGCGACCCGCGCGGCA 93150

Qy        778 GCGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
              ||  | ||   |  |||||||| |||  ||||||| |||||   || || ||| |
Db      93149 CTCGACCAGGGCTTCACCGTCTTCGCCGAGATCAGCCCGCACCCCGTGCTGGTGCCCTCG 93090

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897
              || | |||||  | ||| || || | |    ||     ||| | |||   || |||||
Db      93089 GTCCAGGACACGATCGACGATACGGGCCGGACCGCGGTGGCGGTGGGCTCGCTGCGCCGG 93030

Qy        898 GGCGACAACGACACCCGCCGCTT 920
              | ||||  || | | || |||||
Db      93029 GACGACGGCGGCGCGCGGCGCTT 93007

RESULT 11
AF357202
LOCUS       AF357202              113193 bp    DNA     linear   BCT 17-JUL-2001
```

```
DEFINITION  Streptomyces nodosus amphotericin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF357202
VERSION     AF357202.1  GI:14794889
KEYWORDS    .
SOURCE      Streptomyces nodosus.
  ORGANISM  Streptomyces nodosus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 113193)
  AUTHORS   Caffrey,P., Lynch,S.V., Flood,E.M., Finnan,S.M. and Oliynyk,M.
  TITLE     The amphotericin biosynthetic gene cluster from Streptomyces
            nodosus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 113193)
  AUTHORS   Caffrey,P., Lynch,S.V., Flood,E.M., Finnan,S.M. and Oliynyk,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2001) Industrial Microbiology, University College
            Dublin, Belfield, Dublin, Ireland
FEATURES             Location/Qualifiers
     source          1. .113193
                     /organism="Streptomyces nodosus"
                     /db_xref="taxon:40318"
     gene            complement(4. .1824)
                     /gene="amphG"
     CDS             complement(4. .1824)
                     /gene="amphG"
                     /note="possible amphotericin export protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphG"
                     /protein_id="AAK73498.1"
                     /db_xref="GI:14794890"
                     /translation="MAPDVPEEHEEERESEQPVRRLAALLRPHRRSVGLALTAGVVGI
                     LLNAFGPLLLGRVTDLIADGVLGHGGPAPGVDFGALGRLLMILLVLYVVASVFMLVQN
                     WLVASVVRLLIHDLRHRAQEKLARLPLRYFDRKPAGETLSRGTDDVDNLQQTLQQTLT
                     DLISSVFSLVIMLSLMLIISPSLAGVMLLSVPVSGLLAAWISKRAQPQYAAQWSASGK
                     LTAHVEEMCAGHALVKAFDRRAEAEQRFDERNEAVYRAGSGAQFASGAIEPVMMFVAN
                     LGYVAVAVVGAWKVVNGSLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGVASAQRV
                     FTLLDAPEQEPEPDRPLAVERVEGRVEFQDVSFRYSPDTPLIEGLSLSVEPGSTVAVV
                     GPSGAGKTTVANLLMRFYEIDSGRILLDGTDTAAMNRDDLRSRFGLVLQDTWLFKGTI
                     AENIAYGSPGATRADIVEAARATYADRFIRTLSQGYDTVLDDESGGVSAGEKQLITVA
                     RAFLARPAVLVLDEATSSVDTRTELLIQRAMNTLRPGRTSFVIAHRLSTIRDADVIVV
                     MKSGRIVEQGTHDQLIDAQGAYARLHAARADAPAADVTVG"
     gene            complement(1805. .3628)
                     /gene="amphH"
     CDS             complement(1805. .3628)
                     /gene="amphH"
                     /note="possible amphotericin export protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphH"
                     /protein_id="AAK73499.1"
                     /db_xref="GI:14794891"
                     /translation="MAPSVRWPTPFQRAAAGRRQPGCHSVLLRLMRTQLRPYAGSVVA
                     LVVLHLVQILGTLLLPTLGAALIDEGVVRHDSDRIGTIGATMAVVALVQIAAALGAAA
                     LGARTSTALGRDLRSAVFRRVLDFSAREIGRFGTPSLLTRSVNDVQQVQNLAQSGLGI
                     FVAAPLMCLGSVLLALRQDVTLALILVPMVLVVAVCFGLLLSRMAALYARLQQTLDRI
                     GRLLRERITGVRVVRSFARDAHEGERFTRTNEELLGLSLGVGRLIAVMLPSVLLLMNL
                     FTLGLLWVGARRIDSGSMQIGALSAFLSYLSLILMSVVMLAFVFLNVPRARVCAERIT
                     EVLQAETDVVPPASPRPMAGPAGQVELVGAEFRYPGAENAVLRDLSLTLRPGERVAVL
                     GSTGSGKTTLLHLILRLVDVTAGEVRIGGTDVRELDPSVMAAAVGYVPQRPYLFAGTV
                     ASNMRFGRPDATDEELWEVLRIAQHEGFVTRLGGLDTEIAQGGTTVSGGQRQRLAIAR
                     ALLRRPAIYLFDDSFSALDQSTEAALRKALVPYTEGATVITVAQRVASVRDADRIVLL
                     DQGGIAATGTHDALLRDSPTYREIALSQRTREETAHGAGRS"
     gene            3840. .4874
                     /gene="amphDIII"
     CDS             3840. .4874
                     /gene="amphDIII"
                     /note="mycosamine biosynthesis protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphDIII"
                     /protein_id="AAK73500.1"
                     /db_xref="GI:14794892"
                     /translation="MPKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKFRV
                     SRLASELSFVDGDLMDQGSLVSAVDKVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                     VLRVLEAIRMVSGLSMSRTAGTEGQIRFYQASSSEMFGKVAETPQRETTLFHPRSPYG
                     AAKAYGHFITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKLGLQDKL
                     ALGNMDAVRDWGYAGDYVRAMHLMLQQDAPDDYVIGTGEMHTVRDAVRFAFEHVGLDW
                     KDYVVVDPDLVRPAEVEVLCADSSKAQAQLGWKPSVDFQELMRMMVDADLASVSRQNE
                     LDDLLLAHSW"
     gene            5042. .33574
                     /gene="amphI"
     CDS             5042. .33574
                     /gene="amphI"
                     /note="polyketide synthase multienzyme polypeptide housing
                     extension modules 9, 10, 11, 12, 13, and 14"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphI"
                     /protein_id="AAK73501.1"
                     /db_xref="GI:14794893"
                     /translation="MDNEQKLRDYLKLATADLRRARRRVGELESASQEPIAIVGMTCR
                     YPGGVSSPEDLWRMVEAGENGVTPFPTDRGWDLEALASSPTSRGGFLHDAPEFDADFF
                     GISPREAVAMDPQQRVVLESAWEAFERAGINPTSVKGSRTGVFIGAMAQDYRVGPADG
                     AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                     AGGVTVMSGPGTFIEMGRQGGLSVDGRCRSFGDTADGTGWAEGVGILVLERLSDAIRN
                     GREILAVVRGTAVNQDGASNGLTAPNGPSQQAVIEQALVNARLSAGDIDVVEAHGTGT
                     TLGDPVEAQALLATYGQQRDEDKPLLLGSVKSNISHTQAAAGVAGVIKMVMAMRHGVL
                     PRTLLADEPTRHVDWSQGAVRVLTENTEWPATGAPRRAAVSSFGISGTNAHTIVEQAP
                     EPEPADPEDDAPSTPAAVTGVLPVLLSGRSPEVLRAQAAALLTTLGTGTPPAPADLAY
                     SLATTRTAFEHRAVLLASDLPELTGRLTAIAEGTDPAVLADTVTGTARTETRLAVLFT
                     GQGAQRLGAGRELAARFPVFAAALDAALDAFTPHLDVPLRKVLWGEDADRLDRTEYAQ
                     PALFAVEVALYRLLESFEVKPDHLAGHSVGEIAAAHVAGVFSLDDAATLVAARGRLMQ
                     ALPEGGAMVAVQASEDEVAPLLAGHEDLVSLAAVNGPSAVVLSGDETTVTELAARLAA
                     DGRKTSRLRVSHAFHSPLMAPMLDEFRNVVEGLTLHSPLLPVVSDVTGEPATVAQLTS
                     PDYWVDHVRQAVRFADGIDWLARHDVTAFLELGPDSVLSAMAQNCLDAAGSDALTVPA
                     LREGRPEDHTFTAALAALHTQGTALHWDACFTGTGARRTDLPTYTFQRRRYWPRAVQG
                     GAADLRSVGLGAAHHPLLSAAVSLADSEGALLTGRISLLSHPWLADHTVRGATLLPGT
                     AFLELAVRAGDEVGCDRVDELTLAAPLVLPEQGGVQVQLWIGNPDASGRRSVTVYGRP
                     DADEDAPWTSHATGVLSASRTTSDFDATVWPPADAETLPVDGLYERLAEGGFGYGPLF
                     QGLRAAWRRGDEVFAEVVLPESGHTDAESFGLHPALLDSALHAASFVDLDERAAGGLP
                     FSWEGVSLHASGATTLRVRLAPAAGDAVAIAVADDSGQLVLSADSLILRAVAAREIDA
                     AAALVRDALFRLDWVPVTAVAASGTAAALVGEDPFGLRALPQPGDLAVHPDLADLAAA
                     DGAVPDTVLLPLTGTGPDADPVTAAHRAATEALAAVRTWLEQDERFAASRLALVTRGA
                     TTGHDPAAAAVWGLVRSAQSENPGRFLLVDLDADQDTPALPAAALTSEEPQLAVRGEE
                     LRAARLVRRPASTAEAVPAFGGEGAVLVTGGTGGLGAVLARHLVAEHGVRELVLVSRR
                     GGAAAGAAELVAELAESGARATVVACDVTDRAAVAELVAAHPVSAVVHSAGVLDDGMV
                     GTLTPERLTTVLRPKVDAAWNLHEATRDLDLKAFVLFSSVAGVLGSPGQANYAAGNAF
                     LDALAAHRRAAGLPGLSLAWGPWEQTGGMTGGISEDDLRRMARAGTPALTVEQGLALL
                     DAALDGDDAALAPVRLDLSVLRAQGEVPPLLRSLIRGRSRRAAVAGSATAGGLAQRLA
                     RLDAESRDELVLDLVRGQVALVLGHATGAEIDAGRAFRELGFDSLTAVELRNRLNTVT
                     GLRLPATLVFDYPTVSHLASYVLDELLGTEVEAEVVQRGTAAVADDPIVIVGMACRYP
                     GGVTSPEDLWRLVTEGTDAVSGFPVNRGWDVENLYHPDPDHPGTAYTRSGGFLHEAGE
                     FDPGFFGMSPREALATDSQQRLLLEASWEAIERAGIDPVGLRGSATGVFAGVMYSDYS
                     AMLGSPEFEGFQGSGSSPSLASGRVSYTLGLEGPAVTVDTACSSSLVAMHWAMQALRS
                     GEISLALAGGVTVMSTPAVFVDFARQRGLSPDGRCKAFSDSADGVGWSEGVGMLVLER
                     QSDAIRNGHQILAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALASGGLSAGDVDVV
                     EAHGTGTTLGDPIEAQALLATYGRDRDPEQPLLLGSVKSNIGHTQAAAGVAGVIKMVM
                     SMRHGVLPRTLHVDAPSSHVDWTEGAVELLTEQTAWPETGRPRRAAVSSFGISGTNVH
                     TVLEQAPGTTVPAPAAPERTAGAVPLLLSGRTRDALRAQAARLLTHLQNHPEPSLADL
                     GHSLATTRSRFERRAAVIAQDREGLLASLGSLAAGRPDPAVVEGEAAGRARVAVMFSG
                     QGSQRAAMGRELYETQPRFAAAFDEVCAALDPLLDRPLREVVFAAEGSEEAALLDRTG
                     WTQPALFAVEVALYRLVESWGVRADFVTGHSIGEIAAAHIAGVFTLQDAARLVAARAT
                     LMEALPSGGAMVAVQATEEEVAPLLGEGLSVAAVNGPTSVVVSGDEDPAVELAAEFSG
                     RGRRTKRLRVSHAFHSPHMDAMLDAFRTVAETLSYAAPRIPLVSDLTGRRADDAEVRT
                     ADYWVRHVREAVRFADCVRTLRDAGATLFLELGPDGLLTAMAEDTLGDERYDHNTALV
                     PLLRADRPEESAAATAAARLQIHGVDLDWTAYLAGTGARRVDLPTYAFQHAHYWPQLP
                     SAAPSPAGDPADQKLWAAVERGDAAELAAVLGLDEDSLTPLDSLLPALSSWRRGNQEK
                     ALLDTLRYRVEWTRLSKPAAPVLDGTWLLVSSDATADDETELLDGLAEALGAHGARVR
                     RLVLDADCADRAVLGARLADTENADNTAQVLSVLPLDERPTDGPAGFTQGLALTIALV
                     QALADTGAHGRLWTATRGAVSTGPADPVTHPAQATAWGMGRGVALEHPRLWGGLVDLP
                     ADFDRGAGQRLAEVLAVKDAPDGEDQVALRATGVHGRRLVRHIVDELPSADQFTASGS
                     VLITGGTGGLGAETARWLARSGAAHLVLTSRRGPDAPGAAELRAELEQSGASVSIVAC
                     DVADRDALAAVLDGLSADQPLTGVVHTAGVGHYGPLDALTPAEFAGLTAAKLAGAAHL
```

```
                DNLLGDRELDFFILFGSIAGVWGSGDQSAYGAANAYLDALALARRARGLAATSIAWGP
                WGGTGMAADDAVSGTLRRQGLGLLDPAPALTEMRRAVVRQDVTVTVADVDWTRYAPLF
                TSARPSALISDLPEVRALAAENTPADTGDASEIVQRVRSLSEPEQLRLLTDLVRTEAA
                TVLGHSSAGAVPEDRAFREIGFDSLTAVELRKHLGAATGLSLPSTMVFDYPTPLELAQ
                YLRAEMVGSVLEVAGPVATGGTDDEPIAIIGMSCRYPGGVSSPEQLWDLVLSGTDAIT
                DFPVNRGWNTAGLYDPDPDHPGTTYSTQGGFLHEADEFDPMFFGISPREALVMDPQQR
                LLLETTWEAFERAGLTPDTLRGSLTGTFIGSSYQEYGMGAGDGAEGHLVTGTSPSVLS
                GRLAYVFGLEGPAVTVDTACSSSLVALHLACQALRNGESNLAVAGGATVMTTPNAFVA
                FSRQRALAQDGRCKAFSESADGMTLAEGVGIVLVERLSDARRNGHPVLAVIRGSAINQ
                DGASNGLSAPNGPSQQRVIRQALANARVAPGEIDLLEAHGTGTPLGDPIEAQALFATY
                GRTRTPETALLLGSVKSNIGHSQSAAGVASIIKMVMALRHGVMPQTLHADEPSSHVDW
                SPGTVRLLGENTDWPQTGRPRRAAVSSFGISGTNAHVILEQETEAPAAEDEQLAPAPL
                PVAAGVVPWLLSARGAAALREQADRLLTHLVTADPAARPIDIGLSLATSRALFEHRAV
                VVPPAGTDPLEALRAVAADGPSGVVARGVADVAGRTVFVFPGQGSQWAGMGAQLLDES
                PVFAERIAECAAALAEFTDWNLIDVLRGAEGAPTLERVDVVQPASFAVMVSLAAVWRA
                QGVEPDAVVGHSQGEIAAAVVSGALSLRDGARVVTLRAQAIGRSLAGRGGMMSVALPV
                AEVEARLEAFEGRVSVAAENGPRSSVVAGEPEALDELHAQLTAEEIRARRVAVDYASH
                SPHVEDLHDEILELLAEVAPRTSEIPFFSTVTGDWLDTTVMDAGYWYRSLRGRVLFAD
                AVRDLIAADHRAFIEVSSHPVLAMSVQDMIDDAGVAGVASGTLRRDNGGLDRFLLSAA
                EVFVRGVQVDWAAVFEGTGASRVDLPTYAFQHENLWAMAAAPEAVTAADPEDAAFWTA
                VEDGDVSALTAALGTDEDSVAAVLPALSSWRRARKERSTVDSWRYRPTWKPVTKLPQR
                TLDGTWLLVSADGVDDTDVAEALETGGAEVRRLVLDESCTDRAVLRERLTDADGLTGI
                VSVLAGAERTGAVPGTGLVLGVALTVALVQALGDAGIDTPLWALTRGAVSTGRSDKVT
                APVQAQVTGIGWTAALECPGRWGGVVDLPETLDARAGQRLAAVLAGALGDDDQIALRS
                SGVFTRRIVRADAAPDGSARDWKPRGTTLVTGGSGTLAPHLARWLAEQGAEHLVLVSR
                RGPEAPGAAELRAELAERGTETTLAACDITDRDAVAALLESLKAEGRTVRTVVHTAAT
                IELHTLDATTLDDFDRVLAAKVTGAQILDELLDDEELDDFVLYSSTAGMWGSGAHAAY
                VAGNAYLAALAEHRRARGLTALSLSWGIWADDLQLGRVDPQMIRRSGLEFMDPQLALS
                GLKRALDDDEQVIAVADVDWETYHPVYTSARPTPLFDEVPEVQRLTAAAEQSAGDPAR
                GEFAAALLALPAAEQHRKLLETVRTEAASVLGLSSAEDLTDQRAFRDVGFDSLTAVGL
                RNRLASVTGLTLPSTMVFDYPNPAALAGFLHSELADVHSAGAVAVTAGAPVDDDPIAI
                VGMSCRYPGGITSAEQLWRVSLEEVDAVSVFPADRGWDAEALYDPDPDASGRTYSVQG
                GFLRDVADFDPGFFGISPREALSMDPQQRLLLETAWEVFENAGLDPVAQRGSRTGTFI
                GASYQDYGAAVPGSEGSEGHMITGSLPSVLSGRVSYLFGLEGPAVTLDTACSSSLVAI
                HLACQSLRNGESTLALAGGASIMSTPMSFIGFSRQRALAEDGRCKAYAEGADGMTLAE
                GVGLILLERLSDARRNGHEVLAVIRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAG
                VEANDIDVLEGHGTGTALGDPIEAQALFATYGKDRDPERPVLLGSVKSNIGHTQMASG
                VASIIKLVHALREGVAPKSLHIDQPSTHVDWSSGTIQLLTERTEWPETGRPRRAAVSS
                FGLSGTNVHTVLEQAPAADAPAAEDTPAPRDALVPVLVSGRGEAALRAQAGALLDLLA
                ERPGIHPTDLAFSLATSRAALEHRAAVVADDHEALVRGLTALRDGLPGAGLVQGRTGR
                GRTAFLFTGQGSQRLGMGRELYERHPVFADALDAVLARIDGTTERPLRDVLFAAEGSQ
                DAALLHRTGYAQPALFALEVALFRLLESWGVTPDYLAGHSVGEIAAAHVAGVLDLDDA
                CTLVAARGRLMQALPEGGAMVALEAAEDEVLPHLEGLADQVSVAAVNGPRSVVVAGEE
                EPVLALAAHFAEQGRRTKRLRVSHAFHSPLMDPMLDDFAAVARALTYHAPSIPFVSNV
                TGTLAAPEQVCTADYWVSHVRSAVRFADGIGWLSTQGGVQTFLELGPDGVLSGMARES
                LTDASRTALLPTLRGDRPEEQALVTAVAAAHAHGFDVDWTAWFQGSGARRVALPTYAF
                QRERYWPDTTAAGITAPAPGSALDAEFWAAVEHADVASLTASLGLDDATVTAMVPALT
                AWRQRRGEQSALDSWRYRVTWKPRGGAPGAAPTGRWLVLVPAEHRDEATAAWAADVEA
                ALATATVRVEVTGTDRAALAARLTEAADGDTFQGVLSLLALAPGDAGHPGAPAALTLT
                ATALQALGDARIDAPLWNITRGAVAVGRSEQVTAPEQAAVWGLFRAAALELPARVGGS
                VDLPEDLDTQAARRLRGILAAADGEDAVAVRASGVFLRRLAHSPAADTVGSAFDPAAG
                TVLITGGTGGIGGHLARRLARDGAAHLLLTSRRGPDAPGAGELRAELEESGARVTIAA
                CDAADRDALAALLATVPEDAPLTAVFHTAGVVDDHVVDELTPESFATVLHAKTVAARH

  Query Match          43.5%;  Score 401;  DB 1;  Length 113193;
  Best Local Similarity  64.7%;  Pred. No. 9.9e-40;
  Matches  596;  Conservative    0;  Mismatches  325;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              |||||||| || ||||| |||||     |||||||| || ||||| |  |  || || | |
Db      16592 GTCTTCGTCTTCCCCGGCCAGGGCTCCCAGTGGGCGGGCATGGGCGCCCAGCTCCTGGAC 16651

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
                || ||||||||||||  ||||||    |  ||||  |  ||  ||  ||    ||
Db      16652 GAGTCCCCGGTGTTCGCCGAGCGGATCGCCGAATGCGCCGCCGCCCTCGCCGAGTTCACC 16711

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||    |  | ||| ||||||| ||| |  ||  |||  |   ||  |  |||||  | |
Db      16712 GACTGGAACCTCATCGACGTCCTGCGGGGCGCCGAGGGCGCCCCGACGCTGGAGCGCGTC 16771

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              || || |||||||| |     |||| ||| ||||| ||  |||| ||  |||||||  ||
Db      16772 GACGTCGTCCAGCCCGCCTCGTTCGCCGTGATGGTCTCCCTGGCCGCCGTGTGGCGGGCC 16831

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              | ||| |||||||||||||||  | ||||| || ||||||||||||| ||||||
Db      16832 CAGGGTGTCGAACCCGACGCGGTGGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGGTG 16891

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              || |  || || |||   |||   |||| ||   |   ||  |||| ||| || |||
Db      16892 GTCTCCGGCGCCCTGTCGCTGCGCGACGGCGCCCGGGTCGTGACGCTCCGCGCCCAGGCC 16951

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              | | |       | | | || |||||           | |  ||  ||   | | |
Db      16952 ATCGGCCGCAGCCTCGCCGGGCGCGGCGGCATGATGTCGGTGGCGCTGCCCGTCGCCGAG 17011

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              | | ||  |  |  |||  ||  | ||| |||  |||   ||  ||  ||  |  |||||||||
Db      17012 GTGGAGGCCCGTCTGGAGGCGTTCGAGGGACGGGTGTCCGTCGCCGCCGAGAACGGCCCC 17071

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              ||||||  |||  |||| |||| ||  |||||||  |||||||  |||  ||  |     ||
Db      17072 CGCTCCTCCGTCGTCGCGGGCGAGCCCGAGGCCCTGGACGAACTGCACGCACAGCTCACC 17131

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              | |   | | | |||||||| ||| || | || ||||| ||||||||||  |||||||||
Db      17132 GCCGAGGAGATCCGGGCCCGCCGGGTCGCCGTGGACTACGCCTCGCACTCCCCCCATGTC 17191

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              ||  |||||  ||  ||| | || || |||||||  ||  ||   |||| | ||||||
Db      17192 GAGGACCTGCACGACGAGATCCTCGAACTGCTGGCCGAGGTCGCGCCGCGGACGTCCGAG 17251

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              | ||||||||||| || ||    |||  |||||| ||||||||   | |||||||| |
Db      17252 ATCCCGTTCTTCTCGACCGTCACCGGCGACTGGCTCGACACCACCGTCATGGACGCCGGC 17311

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              ||||||||||||| ||||| |  | ||| |  |||   || || ||||  | ||||
Db      17312 TACTGGTACCGCAGCCTGCGCGGCCGGGTGCTGTTCGCGGACGCAGTCCGTGACCTGATC 17371

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              |  | ||  ||||| | |||| |||| |||||| ||||||||   |||| |    || ||
Db      17372 GCCGCCGACCACCGTGCCTTCATCGAGGTCAGCTCCCACCCCGTGCTCGCCATGTCCGTC 17431

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
              | ||||  | ||| ||| |||  |  |  |   |||   ||||| ||| ||||||| |
Db      17432 CAGGACATGATCGACGACGCCGGAGTGGCGGGTGTCGCCTCCGGCACCCTGCGCCGCGAC 17491

Qy        901 GACAACGACACCCGCCGCTTC 921
              ||  ||    |  |||||||
Db      17492 AACGGCGGTCTCGACCGCTTC 17512


RESULT 12
AX211706
LOCUS       AX211706               27541 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 2 from Patent WO0159126.
ACCESSION   AX211706
VERSION     AX211706.1  GI:15523938
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 27541)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 2 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
```

```
          (NO)
FEATURES             Location/Qualifiers
     source          1. .27541
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT     3517 a  10766 c   9529 g   3729 t
ORIGIN

  Query Match           43.5%;  Score 400.8;  DB 6;  Length 27541;
  Best Local Similarity   66.1%;  Pred. No. 1.5e-39;
  Matches  611;  Conservative    0;  Mismatches  307;  Indels    6;  Gaps    2;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              || |||||||| ||||| ||||||  ||||||||  ||||||||     |  || || |
Db      18020 GTGTTCGTGTTCCCCGGTCAGGGTTCGCAGTGGGTGGGGATGGGGTCCCAACTCCTTGAT 18079

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
                ||  ||||||||||   ||||||    | ||| | || ||| || ||   ||
Db      18080 GAGTCGGCGGTGTTCGCGGAGCGGATTGCCGAGTGTGCGGCGGCACTCGCCGAGTTCACC 18139

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||  |||| ||  | |||||  | |  | ||| |    ||  || |||||||| |
Db      18140 GACTGGTCGCTGGTCGATGTGCTGCGGGGTGTGGTGGGTGCGCCGTCGTTGGAGCGGGTC 18199

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              |||||||| ||||| | |  ||||   || |||||||| ||||||||  ||||||||||
Db      18200 GATGTGGTGCAGCCGGCGTCGTTCGCGGTGATGGTGTCGTTGGCTGCGTTGTGGCGTTCC 18259

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
                ||| |    || || |||||  | ||  ||||| |||| |||||||| || |||
Db      18260 CGTGGTGTGTTGCCGGATGCGGTGGTGGGGCATTCGCAGGGTGAGATCGCTGCTGCGGTG 18319

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||||  || ||||||    ||  ||||| ||||  |   || |||||||| || | |||
Db      18320 GTGTCGGGTGCGCTGTCGTTGCGGGACGGGGCGCGGGTGGTGGCGCTGCGGAGTCAGGCC 18379

Qy        361 CTGG---CCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAG 417
               | |      ||  ||  ||| ||||| || |||    || ||  |||||  | |    | | |
Db      18380 ATTGGTCGTGCGTTGGCGGGGCGGGGAGGGATGATGTCCGTCGCGCTGTCGGTGGACGTG 18439

Qy        418 GTGGAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGC 477
              | || | ||   | |  ||   ||  | ||||||| ||| |||||  || |||||||||
Db      18440 CTCGAACCGCGGTTGGTCGA---GTTCGAGGGGCGGGTGTCGGTGGCCGCCGTCAACGGC 18496

Qy        478 CCCCGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGT 537
              || ||||||  ||  ||| | || ||  |||||||| ||||| | ||| ||
Db      18497 CCGCGCTCCGTCGTGGTCGCCGGCGAGCCCGAGGCGCTGGACGCGCTGCACGCCCGGCTG 18556

Qy        538 GCCGGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCAT 597
              ||| |  ||   | |||||||| |||||| |||| ||||| ||||||||||   | |||
Db      18557 ACCGCCGACGACATCCGGGCCCGCCGGATCGCGGTGGACTACGCCTCGCACTCGCACCAG 18616

Qy        598 GTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCC 657
              || ||  |||||  |||||  |||||||| |||||| |||  |    ||||  ||||
Db      18617 GTCGAGGACCTGCACGAGGAACTGCTGGAGGTGCTGGCGGAGCTGGCGCCGCGCACGTCG 18676

Qy        658 GGCGTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCC 717
              |  |||||||||||||| || |||   |||  ||||||||||||| |     ||||||||
Db      18677 GAGGTGCCGTTCTTCTCGACCGTGACCGGCGACTGGCTGGACACCGCGCGGATGGACGCC 18736

Qy        718 GCCTACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTG 777
              | |||||||| ||||||||||| |    | ||| || |||   || || ||   || ||||
Db      18737 GGCTACTGGTTCCGCAACCTGCGCGGACGGGTGCGGTTCGCGGACGCGGTGGCGGACCTG 18796

Qy        778 GCGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
              ||  |  |  ||||||  |||||||| |||||| | |||||     ||      ||
Db      18797 CTGGCGGCGGAGTACCGCGCATTCGTCGAGGTCAGCTCGCACCCGGTGCTGTCGATGGCG 18856

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897
              |  | ||  | | ||| ||  ||| ||      | |  |||   ||||| ||| ||||||
Db      18857 GTGCAGGAGGCGATCGACGAGGCCGGCGTGCCGGCCGTCGCCGCCGGCACCCTGCGCCGC 18916

Qy        898 GGCGACAACGACACCCGCCGCTTC 921
              | | |   || ||||  |||||||
Db      18917 GACCAGGGCGGCACCGACCGCTTC 18940


RESULT 13
SERERYAB
LOCUS       SERERYAB               20235 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  S.erythraea second and third ORF's of eryA gene, complete cds.
ACCESSION   M63677
VERSION     M63677.1  GI:152693
KEYWORDS    6-deoxyerythronolide B; acyl carrier protein; acyltransferase;
            beta-ketoacyl-ACP synthase I; beta-ketoreductase; dehydratase;
            enoyl reductase; eryA gene.
SOURCE      Saccharopolyspora erythraea DNA.
  ORGANISM  Saccharopolyspora erythraea
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Pseudonocardineae; Pseudonocardiaceae; Saccharopolyspora.
REFERENCE   1  (bases 1 to 20235)
  AUTHORS   Donadio,S., Staver,M.J., McAlpine,J.B., Swanson,S.J. and Katz,L.
  TITLE     Modular organization of genes required for complex polyketide
            biosynthesis
  JOURNAL   Science 252 (5006), 675-679 (1991)
  MEDLINE   91220065
   PUBMED   2024119
FEATURES             Location/Qualifiers
     source          1. .20235
                     /organism="Saccharopolyspora erythraea"
                     /db_xref="taxon:1836"
     gene            19. .20235
                     /gene="eryA"
     CDS             19. .10722
                     /gene="eryA"
                     /note="eryA ORF2 encoding modules 3 & 4 for
                     6-deoxyerythronolide B formation6-deoxyerythronolide B
                     formation; putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26494.1"
                     /db_xref="GI:152694"
                     /translation="MTDSEKVAEYLRRATLDLRAARQRIRELESDPIAIVSMACRLPG
                     GVNTPQRLWELLREGGETLSGFPTDRGWDLARLHHPDPDNPGTSYVDKGGFLDDAAGF
                     DAEFFGVSPREAAAMDPQQRLLLETSWELVENAGIDPHSLRGTATGVFLGVAKFGYGE
                     DTAAAEDVEGYSVTGVAPAVASGRISYTMGLEGPSISVDTACSSSLVALHLAVESLRK
                     GESSMAVVGGAAVMATPGVFVDFSRQRALAADGRSKAFGAGADGFGFSEGVTLVLLER
                     LSEARRNGHEVLAVVRGSALNQDGASNGLSAPSGPAQRRVIRQALESCGLEPGDVDAV
                     EAHGTGTALGDPIEANALLDTYGRDRDADRPLWLGSVKSNIGHTQAAAGVTGLLKVVL
                     ALRNGELPATLHVEEPTPHVDWSSGGVALLAGNQPWRRGERTRRARVSAFGISGTNAH
                     VIVEEAPEREHRETTAHDGRPVPLVVSARTTAALRAQAAQIAELLERPDADLAGVGLG
                     LATTRARHEHRAAVVASTREEAVRGLREIAAGAATADAVVEGVTEVDGRNVVFLFPGQ
                     GSQWAGMGAELLSSSPVFAGKIRACDESMAPMQDWKVSDVLRQAPGAPGLDRVDVVQP
                     VLFAVMVSLAELWRSYGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVVGRSRLMRS
                     LSGEGGMAAVALGEAAVRERLRPWQDRLSVAAVNGPRSVVVSGEPGALRAFSEDCAAE
                     GIRVRDIDVDYASHSPQIERVREELLETTGDIAPRPARVTFHSTVESRSMDGTELDAR
                     YWYRNLRETVRFADAVTRLAESGYDAFIEVSPHPVVVQAVEEAVEEADGAEDAVVVGS
                     LHRDGGDLSAFLRSMATAHVSGVDIRWDVALPGAAPFALPTYPFQRKRYWLQPAAPAA
                     ASDELAYRVSWTPIEKPESGNLDGDWLVVTPLISPEWTEMLCEAINANGGRALRCEVD
                     TSASRTEMAQAVAQAGTGFRGVLSLLSSDESACRPGVPAGAVGLLTLVQALGDAGVDA
                     PVWCLTQGAVRTPADDDLARPAQTTAHGFAQVAGLELPGRWGGVVDLPESVDDAALRL
                     LVAVLRGGGRAEDHLAVRDGRLHGRRVVRASLPQSGSRSWTPHGTVLVTGAASPVGDQ
                     LVRWLADRGAERLVLAGACPGDDLLAAVEEAGASAVVCAQDAAALREALGDEPVTALV
                     HAGTLTNFGSISEVAPEEFAETIAAKTALLAVLDEVLGDRAVEREVYCSSVAGIWGGA
                     GMAAYAAGSAYLDALAEHHRARGRSCTSVAWTPWALPGGAVDDGYLRERGLRSLSADR
                     AMRTWERVLAAGPVSVAVADVDWPVLSEGFAATRPTALFAELAGRGGQAEAEPDSGPT
                     GEPAQRLAGLSPDEQQENLLELVANAVAEVLGHESAAEINVRRAFSELGLDSLNAMAL
                     RKRLSASTGLRLPASLVFDHPTVTALAQHLRARLVGDADQAAVRVVGAADESEPIAIV
                     GIGCRFPGGIGSPEQLWRVLAEGANLTTGFPADRGWDIGRLYHPDPDNPGTSYVDKGG
                     FLTDAADFDPGFFGITPREALAMDPQQRLMLETAWEAVERAGIDPDALRGTDTGVFVG
                     MNGQSYMQLLAGEAERVDGYQGLGNSASVLSGRIAYTFGWEGPALTVDTACSSSLVGI
                     HLAMQALRRGECSLALAGGVTVMSDPYTFVDFSTQRGLASDGRCKAFSARADGFALSE
                     GVAALVLEPLSRARANGHQVLAVLRGSAVNQDGASNGLAAPNGPSQERVIRQALAASG
                     VPAADVDVVEAHGTGTELGDPIEAGALIATYGQDRDRPLRLGSVKTNIGHTQAAAGAA
                     GVIKVVLAMRHGMLPRSLHADELSPHIDWESGAVEVLREEVPWPAGERPRRAGVSSFG
```

```
                     VSGTNAHVIVEEAPAEQEAARTERGPLPFVLSGRSEAVVAAQARALAEHLRDTPELGL
                     TDAAWTLATGRARFDVRAAVLGDDRAGVCAELDALAEGRPSADAVAPVTSAPRKPVLV
                     FPGQGAQWVGMARDLLESSEVFAESMSRCAEALSPHTDWKLLDVVRGDGGPDPHERVD
                     VLQPVLFSIMVSLAELWRAHGVTPAAVVGHSQGEIAAAHVAGALSLEAAAKVVALRSQ
                     VLRELDDQGGMVSVGASRDELETVLARWDGRVAVAAVNGPGTSVVAGPTAELDEFFAE
                     AEAREMKPRRIAVRYASHSPEVARIEDRLAAELGTITAVRGSVPLHSTVTGEVIDTSA
                     MDASYWYRNLRRPVLFEQAVRGLVEQGFDTFVEVSPHPVLLMAVEETAEHAGAEVTCV
                     PTLRRREQSGPHEFLRNLLRAHVHGVGADLRPAVAGGRPAELPTYPFEHQRFWPRPHRP
                     ADVSALGVRGAEHPLLLAAVDVPGHGGAVFTGRLSTDEQPWLAEHVVGGRTLVPGSVL
                     VDLALAAGEDVGLPVLEELVLQRPLVLAGAGALLRMSVGAPDESGRRTIDVHAAEDVA
                     DLADAQWSQHATGTLAQGVAAGPRDTEQWPPEDAVRIPLDDHYDGLAEQGYEYGPSFQ
                     ALRAAWRKDDSVYAEVSIAADEEGYAFHPVLLDAVAQTLSLGALGEPGGGKLPFAWNT
                     VTLHASGATSVRVVATPAGADAMALRVTDPAGHLVATVDSLVVRSTGEKWEQPEPRGG
                     EGELHALDWGRLAEPGSTGRVVAADASDLDAVLRSGEPEPDAVLVRYEPEGDDPRAAA
                     RHGVLWAAALVRRWLEQEELPGATLVIATSGAVTVSDDDSVPEPGAAAMWGVIRCAQA
                     ESPDRFVLLDTDAEPGMLPAVPDNPQLALRGDDVFVPRLSPLAPSALTLPAGTQRLVP
                     GDGAIDSVAFEPAPDVEQPLRAGEVRVDVRATGVNFRDVLLALGMYPQKADMGTEAAG
                     VVTAVGPDVDAFAPGDRVLGLFQGAFAPIAVTDHRLLARVPDGWSDADAAAVPIAYTT
                     AHYALHDLAGLRAGQSVLIHAAAGGVGMAAVALARRAGAEVLATAGPAKHGTLRALGL
                     DDEHIASSRETGFARKFRERTGGRGVDVVLNSLTGELLDESADLLAEDGVFVEMGKTD
                     LRDAGDFRGRYAPFDLGEAGDDRLGEILREVVGLLGAGELDRLPVSAWELGSAPAALQ
                     HMSRGRHVGKLVLTQPAPVDPDGTVLITGGTGTLGRLLARHLVTEHGVRHLLLVSRRG
                     ADAPGSDELRAEIEDLGASAEIAACDTADRDALSALLDGLPRPLTGVVHAAGVLADGL
                     VTSIDEPAVEQVLRAKVDAAWNLHELTANTGLSFFVLFSSAASVLAGPGQGVYAAANE
                     SLNALAALRRTRGLPAKALGWGLWAQASEMTSGLGDRIARTGVAALPTERALALFDSA
                     LRRGGEVVFPLSINRSALRRAEFVPEVLRGMVRAKLRAAGQAEAAGPNVVDRLAGRSE
                     SDQVAGLAELVRSHAAAVSGYGSADQLPERKAFKDLGFDSLAAVELRNRLGTATGVRL
                     PSTLVFDHPTPLAVAEHLRDRLFAASPAVDIGDRLDELEKALEALSAEDGHDDVGQRL
                     ESLLRRWNSRRADAPSTSAISEDASDDELFSMLDQRFGGGEDL"
misc_feature         19. .4470
                     /gene="eryA"
                     /note="putative"
                     /function="'approximate span of module 3'"
misc_feature         97. .1482
                     /gene="eryA"
                     /note="putative"
                     /function="'approximate span of beta-ketoacyl ACP synthase
                     of module 3'"
misc_feature         1693. .2670
                     /gene="eryA"
                     /note="putative"
                     /function="'approximate span of acyltransferase domain of
                     module 3'"
misc_feature         3406. .3921
                     /gene="eryA"
                     /note="putative"
                     /function="'approximate span of beta-ketoreductase domain
                     of module 3, possibly non-functional'"
misc_feature         4171. .4428
                     /gene="eryA"
                     /note="putative"
                     /function="'approximate span of acyl carrier domain of
                     module 3'"
misc_feature         4471. .10722
                     /gene="eryA"
                     /note="putative"
                     /function="'approximate span of module 4'"
misc_feature         4471. .5847
                     /gene="eryA"
                     /note="putative"
                     /function="'approximate span of beta-ketoacyl ACP synthase
                     domain of module 4'"
misc_feature         6054. .7026
                     /gene="eryA"
                     /note="putative"
                     /function="'approximate span of acyltransferase domain of
                     module 4'"
misc_feature         7165. .9216
                     /gene="eryA"
                     /note="putative"
                     /function="'approximate span of dehydratase and
                     enoylreductase domains of module 4'"
misc_feature         9433. .9984
                     /gene="eryA"
                     /note="putative"
                     /function="'approximate span of beta-ketoreductase domain
                     of module 4'"
misc_feature         10225. .10483
                     /gene="eryA"
                     /note="putative"
                     /function="'approximate span of acyl carrier domain of
                     module 4'"
CDS                  10723. .20235
                     /gene="eryA"
                     /note="ORF3 encoding modules 5 & 6 for
                     6-deoxyerythronolide B formation; putative"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAA26495.1"
                     /db_xref="GI:152695"
                     /translation="MSGDNGMTEEKLRRYLKRTVTELDSVTARLREVEHRAGEPIAIV
                     GMACRFPGDVDSPESFWEFVSGGGDAIAEAPADRGWEPDPDARLGGMLAAAGDFDAGF
                     FGISPREALAMDPQQRIMLEISWEALERAGHDPVSLRGSATGVFTGVGTVDYGPRPDE
                     APDEVLGYVGTGTASSVASGRVAYCLGLEGPAMTVDTACSSGLTALHLAMESLRRDEC
                     GLALAGGVTVMSSPGAFTEFRSQGGLAADGRCKPFSKAADGFGLAEGAGVLVLQRLSA
                     ARREGRPVLAVLRGSAVNQDGASNGLTAPSGPAQQRVIRRALENAGVRAGDVDYVEAH
                     GTGTRLGDPIEVHALLSTYGAERDPDDPLWIGSVKSNIGHTQAAAGVAGVMKAVLALR
                     HGEMPRTLHFDEPSPQIEWDLGAVSVVSQARSWPAGERPRRAGVSSFGISGTNAHVIV
                     EEAPEADEPEPAPDSGPVPLVLSGRDEQAMRAQAGRLADHLAREPRNSLRDTGFTLAT
                     RRSAWEHRAVVVGDRDDALAGLRAVADGRIADRTATGQARTRRGVAMVFPGQGAQWQG
                     MARDLLRESQVFADSIRDCERALAPHVDWSLTDLLSGARPLDRVDVVQPALFAVMVSL
                     AALWRSHGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVAVRSRVLARLGGQGGMAS
                     FGLGTEQAAERIGRFAGALSIASVNGPRSVVVAGESGPLDELIAECEAEGITARRIPV
                     DYASHSPQVESLREELLTELAGISPVSADVALYSTTTGQPIDTATMDTAYWYANLREQ
                     VRFQDATRQLAEAGFDAFVEVSPHPVLTVGIEATLDSALPADAGACVVGTLRRDRGGL
                     ADFHTALGEAYAQGVEVDWSPAFADARPVELPVYPFQRQRYWLPIPTGGRARDEDDDW
                     RYQVVWREAEWESASLAGRVLLVTGPGVPSELSDAIRSGLEQSGATVLTCDVESRSTI
                     GTALEAADTDALSTVVSLLSRDGEAVDPSLDALALVQALGAAGVEAPLWVLTRNAVQV
                     ADGELVDPAQAMVGGLGRVVGIEQPGRWGGLVDLVDADAASIRSLAAVLADPRGEEQV
                     AIRADGIKVARLVPAPARAARTRWSPRGTVLVTGGTGGIGAHVARWLARSGAEHLVLL
                     GRRGADAPGASELREELTALGTGVTIAACDVADRARLEAVLAAERAEGRTVSAVMHAA
                     GVSTSTPLDDLTEAEFTEIADVKVRGTVNLDELCPDLDAFVLFSSNAGVWGSPGLASY
                     AAANAFLDGFARRRRSEGAPVTSIAWGLWAGQNMAGDEGGEYLRSQGLRAMDPDRAVE
                     ELHITLDHGQTSVSVVDMDRRRFVELFTAARHRPLFDEIAGARAEARQSEEGPALAQR
                     LAALSTAERREHLAHLIRAEVAAVLGHGDDAAIDRDRAFRDLGFDSMTAVDLRNRLAA
                     VTGVREAATVVFDHPTITRLADHYLERLVGAAEAEQAPALVREVPKDADDPIAIVGMA
                     CRFPGGVHNPGELWEFIVGRGDAVTEMPTDRGWDLDALFDPDPQRHGTSYSRHGAFLD
                     GAADFDAAFFGISPREALAMDPQQRQVLETTWELFENAGIDPHSLRGSDTGVFLGAAY
                     QGYGQDAVVPEDSEGYLLTGNSSAVVSGRVAYVLGLEGPAVTVDTACSSSLVALHSAC
                     GSLRDGDCGLAVAGGVSVMAGPEVFTEFSRQGGLAVDGRCKAFSAEADGFGFAEGVAV
                     VLLQRLSDARRAGRQVLGVVAGSAINQDGASNGLAAPSGVAQQRVIRKAWARAGITGA
                     DVAVVEAHGTGTRLGDPVEASALLATYGKSRGSSGPVLLGSVKSNIGHAQAAAGVAGV
                     IKVVLGLNRGLVPPMLCRGERSPLIEWSSGGVELAEAVSPWPPAADGVRRAGVSAFGV
                     SGTNAHVIIAEPPEPEPLPEPGPVGVLAAANSVPVLLSARTETALAAQARLLESAVDD
                     SVPLTALASALATGRAHLPRRAALLAGDHEQLRGQLRAVAEGVAAPGATTGTASAGGV
                     VFVFPGQGAQWEGMARGLLSVPVFAESIAECDAVLSEVAGFSASEVLEQRPDAPSLER
                     VDVVQPVLFSVMVSLARLWGACGVSPSAVIGHSQGEIAAAVVAGVLSLEDGVRVVALR
                     AKALRALAGKGGMVSLAAPGERARALIAPWEDRISVAAVNSPSSVVVSGDPEALAELV
                     ARCEDEGVRAKTLPVDYASHSRHVEEIRETILADLDGISARRAAIPLYSTLHGERRDG
                     ADMGPRYWYDNLRSQVRFDEAVSAAVADGHATFVEMSPHPVLTAAVQEIAADAVAIGS
                     LHRDTAEEHLIAELARAHVHGVAVDWRNVFPAAPPVALPNYPFEPQRYWLAPEVSDQL
                     ADSRYRVDWRPLATTPVDLEGGFLVHGSAPESLTSAVEKAGGRVVPVASADREASAAL
                     REVPGEVAGVLSVHTGAATHLALHQSLGEAGVRAPLWLVTSRAVALGESEPVDPEQAM
                     VWGLGRVMGLETPERWGGLVDLPAEPAPGDGEAFVACLGADGHEDQVAIRDHARYGRR

  Query Match             41.8%;  Score 385.4;  DB 1;  Length 20235;
  Best Local Similarity   65.6%;  Pred. No. 1.1e-37;
  Matches  579;  Conservative    0;  Mismatches  301;  Indels    3;  Gaps    1;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              |||||| |||| || || |||||| ||| |||||||| ||||||     |||||  |
Db       1696 GTCTTCCTGTTCCCGGGGCAGGGTTCGCAATGGGCCGGCATGGGTGCCGAGCTGCTGTCG 1755

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              || || |||||||||||| ||  ||| | ||| ||||| ||| |  |||| ||| |  |
Db       1756 TCGTCGCCGGTGTTCGCCGGGAAGATCCGGGCCTGCGACGAGTCGATGGCCCCGATGCAG 1815
```

```
Qy       121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
Db      1816 GACTGGAAGGTCTCCGACGTGCTGCGTCAGGCGCCGGGGGCGCCGGGCCTGGACCGGGTC 1875

Qy       181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
Db      1876 GACGTGGTGCAGCCGGTGTTGTTCGCGGTGATGGTGTCGCTGGCGGAGCTGTGGCGCTCG 1935

Qy       241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
Db      1936 TACGGCGTGGAGCCCGCGGCGGTCGTGGGGCACTCGCAGGGCGAGATCGCCGCCGCGCAC 1995

Qy       301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
Db      1996 GTCGCCGGGGCGCTCACGTTGGAGGACGCGGCGAAGCTCGTCGTGGGCCGCAGCCGCCTG 2055

Qy       361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
Db      2056 ATGCGGTCGCTCTCCGGGGAGGGCGGCATGGCCGCCGTCGCGCTGGGCGAGGCCGCGGTG 2115

Qy       421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
Db      2116 CGCGAGCGCCTGCG---GCCGTGGCAGGACCGGCTCTCGGTGGCCGCGGTCAACGGTCCC 2172

Qy       481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
Db      2173 CGGTCGGTCGTGGTCTCCGGCGAGCCCGGCGCGCTGCGGGCGTTTTCCGAGGACTGCGCG 2232

Qy       541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
Db      2233 GCCGAGGGCATCCGCGTCCGCGACATCGACGTGGACTACGCCTCGCACTCGCCGCAGATC 2292

Qy       601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
Db      2293 GAGCGGGTCCGCGAGGAACTCCTCGAAACGACCGGCGACATCGCGCCGCGCCCGGCGCGG 2352

Qy       661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
Db      2353 GTGACGTTCCACTCCACTGTGGAGTCGCGGTCTATGGACGGCACCGAGCTGGATGCCCGG 2412

Qy       721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
Db      2413 TACTGGTACCGCAACCTGCGCGAGACGGTGCGCTTCGCCGACGCCGTGACGCGGCTGGCG 2472

Qy       781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
Db      2473 GAGTCGGGATACGACGCGTTCATCGAGGTCAGCCCGCATCCGGTCGTGGTCCAGGCCGTC 2532

Qy       841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCG 883
Db      2533 GAGGAGGCGGTCGAAGAGGCTGACGGTGCCGAAGACGCGGTCG 2575
```

RESULT 14
AR049368
LOCUS AR049368 20235 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 3 from patent US 5824513.
ACCESSION AR049368
VERSION AR049368.1 GI:6005407
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 20235)
AUTHORS Katz,L., Donadio,S. and McAlpine,J.B.
TITLE Recombinant DNA method for producing erythromycin analogs
JOURNAL Patent: US 5824513-A 3 20-OCT-1998;
FEATURES Location/Qualifiers
source 1. .20235
/organism="unknown"
BASE COUNT 2534 a 7194 c 7767 g 2740 t
ORIGIN

Query Match 41.8%; Score 385.4; DB 6; Length 20235;
Best Local Similarity 65.6%; Pred. No. 1.1e-37;
Matches 579; Conservative 0; Mismatches 301; Indels 3; Gaps 1;

```
Qy         1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
Db      1696 GTCTTCCTGTTCCCGGGGCAGGGTTCGCAATGGGCCGGCATGGGTGCCGAGCTGCTGTCG 1755

Qy        61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
Db      1756 TCGTCGCCGGTGTTCGCCGGGAAGATCCGGGCCTGCGACGAGTCGATGGCCCCGATGCAG 1815

Qy       121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
Db      1816 GACTGGAAGGTCTCCGACGTGCTGCGTCAGGCGCCGGGGGCGCCGGGCCTGGACCGGGTC 1875

Qy       181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
Db      1876 GACGTGGTGCAGCCGGTGTTGTTCGCGGTGATGGTGTCGCTGGCGGAGCTGTGGCGCTCG 1935

Qy       241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
Db      1936 TACGGCGTGGAGCCCGCGGCGGTCGTGGGGCACTCGCAGGGCGAGATCGCCGCCGCGCAC 1995

Qy       301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
Db      1996 GTCGCCGGGGCGCTCACGTTGGAGGACGCGGCGAAGCTCGTCGTGGGCCGCAGCCGCCTG 2055

Qy       361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
Db      2056 ATGCGGTCGCTCTCCGGGGAGGGCGGCATGGCCGCCGTCGCGCTGGGCGAGGCCGCGGTG 2115

Qy       421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
Db      2116 CGCGAGCGCCTGCG---GCCGTGGCAGGACCGGCTCTCGGTGGCCGCGGTCAACGGTCCC 2172

Qy       481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
Db      2173 CGGTCGGTCGTGGTCTCCGGCGAGCCCGGCGCGCTGCGGGCGTTTTCCGAGGACTGCGCG 2232

Qy       541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
Db      2233 GCCGAGGGCATCCGCGTCCGCGACATCGACGTGGACTACGCCTCGCACTCGCCGCAGATC 2292

Qy       601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
Db      2293 GAGCGGGTCCGCGAGGAACTCCTCGAAACGACCGGCGACATCGCGCCGCGCCCGGCGCGG 2352

Qy       661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
Db      2353 GTGACGTTCCACTCCACTGTGGAGTCGCGGTCTATGGACGGCACCGAGCTGGATGCCCGG 2412

Qy       721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
Db      2413 TACTGGTACCGCAACCTGCGCGAGACGGTGCGCTTCGCCGACGCCGTGACGCGGCTGGCG 2472

Qy       781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
Db      2473 GAGTCGGGATACGACGCGTTCATCGAGGTCAGCCCGCATCCGGTCGTGGTCCAGGCCGTC 2532

Qy       841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCG 883
Db      2533 GAGGAGGCGGTCGAAGAGGCTGACGGTGCCGAAGACGCGGTCG 2575
```

RESULT 15
AR095529
LOCUS AR095529 20235 bp DNA linear PAT 08-SEP-2000
DEFINITION Sequence 3 from patent US 6004787.
ACCESSION AR095529
VERSION AR095529.1 GI:10023473
KEYWORDS .

```
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 20235)
  AUTHORS   Katz,L., Donadio,S. and McAlpine,J.B.
  TITLE     Method of directing biosynthesis of specific polyketides
  JOURNAL   Patent: US 6004787-A 3 21-DEC-1999;
FEATURES             Location/Qualifiers
     source          1. .20235
                     /organism="unknown"
BASE COUNT     2534 a   7194 c   7767 g   2740 t
ORIGIN
```

Query Match 41.8%; Score 385.4; DB 6; Length 20235;
Best Local Similarity 65.6%; Pred. No. 1.1e-37;
Matches 579; Conservative 0; Mismatches 301; Indels 3; Gaps 1;

```
Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              |||||| |||| || || ||||||  ||| |||||||| |||||||      |||||  |
Db       1696 GTCTTCCTGTTCCCGGGGCAGGGTTCGCAATGGGCCGGCATGGGTGCCGAGCTGCTGTCG 1755

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              || || |||||||||||| ||  ||| | ||| ||||| ||| |  |||| |||  |  |
Db       1756 TCGTCGCCGGTGTTCGCCGGGAAGATCCGGGCCTGCGACGAGTCGATGGCCCCGATGCAG 1815

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||   ||    ||| | ||||| | ||   ||||||   |||     ||| |||| |
Db       1816 GACTGGAAGGTCTCCGACGTGCTGCGTCAGGCGCCGGGGGCGCCGGGCCTGGACCGGGTC 1875

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              || ||||| ||||| ||| |||||   || ||||||||  |||| |  |||||||| ||
Db       1876 GACGTGGTGCAGCCGGTGTTGTTCGCGGTGATGGTGTCGCTGGCGGAGCTGTGGCGCTCG 1935

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              |||||  | || ||||  |||||| | || || || |||||||||||||| ||||||||
Db       1936 TACGGCGTGGAGCCCGCGGCGGTCGTGGGGCACTCGCAGGGCGAGATCGCCGCCGCGCAC 1995

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||    |||||||| |   || ||||||||||||||   || | |   ||||||||    |
Db       1996 GTCGCCGGGGCGCTCACGTTGGAGGACGCGGCGAAGCTCGTCGTGGGCCGCAGCCGCCTG 2055

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              ||   |  |   ||  |||||||||||| | ||  |||||   |     | ||||
Db       2056 ATGCGGTCGCTCTCCGGGGAGGGCGGCATGGCCGCCGTCGCGCTGGGCGAGGCCGCGGTG 2115

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              |||  | |   |   || ||||  ||  ||| | | |||||| ||||||||||| |||
Db       2116 CGCGAGCGCCTGCG---GCCGTGGCAGGACCGGCTCTCGGTGGCCGCGGTCAACGGTCCC 2172

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              || ||   ||  ||||| || ||  |||  ||| ||   | | |     | | |||| ||
Db       2173 CGGTCGGTCGTGGTCTCCGGCGAGCCCGGCGCGCTGCGGGCGTTTTCCGAGGACTGCGCG 2232

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              | |   ||  | || | |||    |||   || ||||| ||||||||||  || ||  |
Db       2233 GCCGAGGGCATCCGCGTCCGCGACATCGACGTGGACTACGCCTCGCACTCGCCGCAGATC 2292

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              |||   | || |||||  | || ||   |   || ||||||  ||||  ||| |  |
Db       2293 GAGCGGGTCCGCGAGGAACTCCTCGAAACGACCGGCGACATCGCGCCGCGCCCGGCGCGG 2352

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              ||| |||||  |||||| ||||||       | ||||  |||    |||||  |||
Db       2353 GTGACGTTCCACTCCACTGTGGAGTCGCGGTCTATGGACGGCACCGAGCTGGATGCCCGG 2412

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              |||||||||||||||||| | || |||| || |||  |||  |||||   |   ||||||
Db       2413 TACTGGTACCGCAACCTGCGCGAGACGGTGCGCTTCGCCGACGCCGTGACGCGGCTGGCG 2472

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              ||    ||| ||  ||  ||| |||| |||||||| || ||   | | ||||  ||| ||
Db       2473 GAGTCGGGATACGACGCGTTCATCGAGGTCAGCCCGCATCCGGTCGTGGTCCAGGCCGTC 2532
```

```
Qy         841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCG 883
               || ||  |   ||||||  | | ||  | ||    |||| |||
Db        2533 GAGGAGGCGGTCGAAGAGGCTGACGGTGCCGAAGACGCGGTCG 2575
```

Search completed: June 17, 2003, 17:55:32
Job time : 1791.01 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 183.212 Seconds
(without alignments)
11320.736 Million cell updates/sec

Title: US-09-914-286-1_COPY_3148_4068
Perfect score: 921
Sequence: 1 gtcttcgtgtttcccggaca.....acaacgacacccgccgcttc 921

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	921	100.0	11916	22	AAH79279	Streptomyces averm
	2	921	100.0	12381	21	AAZ58381	Streptomyces averm
	3	921	100.0	30690	21	AAA92301	S. avermitilis ave
	4	921	100.0	30690	22	AAH79277	Streptomyces averm
	5	919.4	99.8	31422	21	AAA92302	S. avermitilis ave
	6	919.4	99.8	31422	22	AAH79278	Streptomyces averm
	7	402.4	43.7	65140	22	AAD17184	Streptomyces nours
	8	402.4	43.7	125401	22	AAD17186	Streptomyces nours
	9	400.8	43.5	27541	22	AAD17185	Streptomyces nours
	10	385.4	41.8	29879	14	AAQ46806	eryA region of S.
	11	373.8	40.6	50937	21	AAA09469	Streptococcus olea
	12	362.8	39.4	1010	20	AAV99255	DNA encoding an ac
	13	362.8	39.4	1010	21	AAA72521	DNA sequence of th
	14	351	38.1	53789	19	AAV21187	Amycolatopsis medi
	15	349.6	38.0	7788	22	AAF88335	S. spinosa DNA fra
	16	349.6	38.0	50000	22	AAF88312	S. spinosa DNA fra
	17	349.6	38.0	50000	22	AAF88316	S. spinosa DNA fra
	18	348	37.8	80161	20	AAZ21501	DNA fragment of Sa
	19	347.4	37.7	47981	22	AAF30757	Micromonospora meg
	20	345	37.5	28598	17	AAT06769	Sorangium cellulos
	21	345	37.5	28958	18	AAT89956	Sorangium cellulos
	22	345	37.5	28958	21	AAA75299	DNA sequence of So
	23	345	37.5	49377	19	AAV05287	The soraphen biosy
	24	344.4	37.4	44377	18	AAT78508	Platenolide syntha
	25	344.4	37.4	44377	18	AAT80414	Platenolide syntha
	26	342.4	37.2	9513	22	AAF88337	S. spinosa DNA fra
	27	342.4	37.2	50000	22	AAF88313	S. spinosa DNA fra
	28	342.2	37.2	43280	18	AAT80413	Tylactone synthase
	29	340.4	37.0	16767	22	AAF88339	S. spinosa DNA fra
	30	319	34.6	4411529	22	AAI99682	Mycobacterium tube
	31	318	34.5	13842	21	AAZ87297	S. venezuelae macr
	32	318	34.5	36778	21	AAZ87318	S. venezuelae pik
	33	318	34.5	37948	21	AAZ87285	S. venezuelae pik
	34	318	34.5	38506	21	AAA75633	Nucleotide sequenc
	35	318	34.5	38506	21	AAZ56001	Recombinant cosmid
c	36	317.2	34.4	77536	21	AAA14651	Nucleotide sequenc
	37	311	33.8	11220	21	AAZ87298	S. venezuelae macr
	38	308.6	33.5	1035	20	AAV99256	DNA encoding an ac
	39	308.6	33.5	1035	21	AAA72538	DNA encoding the A
	40	308.6	33.5	4403765	22	AAI99683	Mycobacterium tube
c	41	306.8	33.3	4403765	22	AAI99683	Mycobacterium tube
c	42	306.8	33.3	4411529	22	AAI99682	Mycobacterium tube
	43	305.8	33.2	4689	21	AAZ87299	S. venezuelae macr
c	44	302	32.8	1881	24	AAS18436	Contig 109 DNA enc
	45	298.4	32.4	68750	21	AAZ55887	Sorangium cellulos

ALIGNMENTS

RESULT 1
AAH79279
ID AAH79279 standard; DNA; 11916 BP.
XX
AC AAH79279;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis coding sequence derivative SEQ ID NO: 3.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..11916
FT /*tag= a
FT /product= "AAG65268"
FT /partial
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX

```
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65268.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 10; Page 149-167; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is modified version of a
CC   fragment of the S. avermitilis genome.
XX
SQ   Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;

  Query Match             100.0%;  Score 921;  DB 22;  Length 11916;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-146;
  Matches  921;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3148 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 3207

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3208 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 3267

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3268 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 3327

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3328 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 3387

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3388 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 3447

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3448 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 3507

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3508 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 3567

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3568 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 3627

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3628 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 3687

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3688 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 3747

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3748 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 3807

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3808 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 3867

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3868 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 3927

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3928 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 3987

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3988 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 4047

Qy        901 GACAACGACACCCGCCGCTTC 921
              |||||||||||||||||||||
Db       4048 GACAACGACACCCGCCGCTTC 4068


RESULT 2
AAZ58381
ID   AAZ58381 standard; DNA; 12381 BP.
XX
AC   AAZ58381;
XX
DT   23-MAY-2000  (first entry)
XX
DE   Streptomyces avermitilis avermectin polyketide synthase modules 1+2.
XX
KW   Polyketide synthase; avermectin; insecticide; ss.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200001827-A2.
XX
PD   13-JAN-2000.
XX
PF   06-JUL-1999;   99WO-GB02158.
XX
PR   06-JUL-1998;   98GB-0014622.
XX
PA   (BIOT-) BIOTICA TECHNOLOGY LTD.
PA   (PFIZ ) PFIZER INC.
XX
PI   Kellenberger JL,  Leadlay PF,  Staunton J,  Stutzman-Engwall KJ;
PI   McArthur HAI;
XX
DR   WPI; 2000-182117/16.
XX
PT   Mutated Type I polyketide synthase containing a polylinker site in an
PT   extension module for replacement of a reductive loop sequence, for
PT   producing polyketides, e.g. B1 avermectin -
XX
PS   Disclosure; Fig 7a-f; 75pp; English.
XX
CC   The present sequence is that of DNA encoding the first 2 modules
CC   of the avermectin polyketide synthase (PKS) of Streptomyces
CC   avermitilis.  The invention relates to nucleic acids encoding a
CC   Type I PKS such as avermectin in which a polylinker with multiple
CC   restriction sites replaces or 1 more PKS genes encoding enzymes
CC   associated with reduction.  Novel PKS are provided in which in
CC   which the reductive loop in a selected module of the Type I PKS is
CC   replaced with the equivalent segment from the same or different
CC   PKS gene cluster or by a mutated or synthetic segment.  Vectors and
CC   host cells, and methods for producing novel polyketides by
CC   culturing host cells are claimed.  The polyketides obtained are
CC   useful as antibiotics and insecticides.  Fermentation products
CC   containing C22-C23 dihydroavermectin, ivermectin and B1
CC   avermectins are claimed.
XX
SQ   Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;

  Query Match             100.0%;  Score 921;  DB 21;  Length 12381;
```

```
  Best Local Similarity   100.0%;  Pred. No. 1.6e-146;
  Matches  921;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3965 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 4024

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4025 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 4084

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4085 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 4144

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4145 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 4204

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4205 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 4264

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4265 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 4324

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4325 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 4384

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4385 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 4444

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4445 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 4504

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4505 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 4564

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4565 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 4624

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4625 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 4684

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4685 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 4744

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4745 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 4804

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4805 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 4864

Qy        901 GACAACGACACCCGCCGCTTC 921
              |||||||||||||||||||||
Db       4865 GACAACGACACCCGCCGCTTC 4885


RESULT 3
AAA92301
ID   AAA92301 standard; DNA; 30690 BP.
XX
AC   AAA92301;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag=  a
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             11971..30690
FT                   /*tag=  b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23749, AAB23750.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 2; Page 66-134; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;

  Query Match           100.0%;  Score 921;  DB 21;  Length 30690;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-146;
  Matches  921;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3148 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 3207

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3208 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 3267

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3268 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 3327

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3328 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 3387
```

```
Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3388 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 3447

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3448 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 3507

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3508 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 3567

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3568 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 3627

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3628 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 3687

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3688 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 3747

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3748 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 3807

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3808 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 3867

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3868 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 3927

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3928 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 3987

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3988 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 4047

Qy        901 GACAACGACACCCGCCGCTTC 921
              |||||||||||||||||||||
Db       4048 GACAACGACACCCGCCGCTTC 4068
```

```
RESULT 4
AAH79277
ID   AAH79277 standard; DNA; 30690 BP.
XX
AC   AAH79277;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 1.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag=  a
FT                   /product= "AAG65264"
FT                   /partial
FT   CDS             11971..30690
FT                   /*tag=  b
FT                   /product= "AAG65265"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65264, AAG65265.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Example 2; Page 58-123; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;
```

```
  Query Match           100.0%;  Score 921;  DB 22;  Length 30690;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-146;
  Matches  921;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3148 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 3207

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3208 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 3267

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3268 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 3327

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3328 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 3387

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3388 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 3447

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3448 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 3507

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3508 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 3567

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3568 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 3627

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       3628 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 3687

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3688 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 3747

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3748 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 3807

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3808 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 3867

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3868 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 3927

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3928 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 3987

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3988 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 4047

Qy        901 GACAACGACACCCGCCGCTTC 921
              |||||||||||||||||||||
Db       4048 GACAACGACACCCGCCGCTTC 4068
```

```
RESULT 5
AAA92302
ID   AAA92302 standard; DNA; 31422 BP.
XX
AC   AAA92302;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag= a
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             14824..31422
FT                   /*tag= b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23751, AAB23752.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use -
XX
PS   Claim 2; Page 134-203; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;

  Query Match             99.8%;  Score 919.4;  DB 21;  Length 31422;
  Best Local Similarity   99.9%;  Pred. No. 2.9e-146;
  Matches  920;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16630 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 16689

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16690 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 16749

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16750 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 16809

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16810 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 16869

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16870 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 16929

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16930 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 16989

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16990 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 17049

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17050 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 17109

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17110 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 17169

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17170 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 17229

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17230 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 17289

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17290 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 17349

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17350 TACTGGTACCGCAACCTGCACCAGCCTGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 17409
```

```
Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17410 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 17469

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17470 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 17529

Qy        901 GACAACGACACCCGCCGCTTC 921
              |||||||||||||||||||||
Db      17530 GACAACGACACCCGCCGCTTC 17550


RESULT 6
AAH79278
ID   AAH79278 standard; DNA; 31422 BP.
XX
AC   AAH79278;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 2.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag=  a
FT                   /product= "AAG65266"
FT   CDS             14824..31422
FT                   /*tag=  b
FT                   /product= "AAG65267"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65266, AAG65267.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Disclosure; Page 103-149; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;

  Query Match             99.8%;  Score 919.4;  DB 22;  Length 31422;
  Best Local Similarity   99.9%;  Pred. No. 2.9e-146;
  Matches  920;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16630 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 16689

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16690 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 16749

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16750 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 16809

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16810 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 16869

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16870 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 16929

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16930 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 16989

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16990 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 17049

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17050 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 17109

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17110 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 17169

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17170 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 17229

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17230 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 17289

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17290 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 17349

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      17350 TACTGGTACCGCAACCTGCACCAGCCTGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 17409

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17410 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 17469

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      17470 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 17529

Qy        901 GACAACGACACCCGCCGCTTC 921
              |||||||||||||||||||||
Db      17530 GACAACGACACCCGCCGCTTC 17550


RESULT 7
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
```

```
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..1035)
FT                   /*tag= a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement (1056..2576)
FT                   /*tag= b
FT                   /product= "NysD1 protein"
FT   CDS             2806..6906
FT                   /*tag= c
FT                   /product= "NysA protein"
FT   CDS             6952..16530
FT                   /*tag= d
FT                   /product= "NysB protein"
FT   CDS             16550..49840
FT                   /*tag= e
FT                   /product= "NysC protein"
FT   CDS             50260..51015
FT                   /*tag= f
FT                   /product= "NysE protein"
FT   CDS             51405..54305
FT                   /*tag= g
FT                   /product= "NysR1 protein"
FT   CDS             54329..57190
FT                   /*tag= h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180..59963
FT                   /*tag= i
FT                   /product= "NysR3 protein"
FT   CDS             60415..61047
FT                   /*tag= j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736..62497
FT                   /*tag= k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (62551..63615)
FT                   /*tag= l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765..64961
FT                   /*tag= m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI   Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 116-151; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

  Query Match           43.7%;  Score 402.4;  DB 22;  Length 65140;
  Best Local Similarity  66.2%;  Pred. No. 2.3e-59;
  Matches  612;  Conservative    0;  Mismatches  306;  Indels    6;  Gaps    2;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              || |||||||| ||||| ||||||  ||||||||  ||||||||  |   |  || ||||
Db       8674 GTGTTCGTGTTCCCCGGTCAGGGTTCGCAGTGGGTGGGGATGGGGGCCCAACTCCTCGAT 8733

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
                ||  ||||||||||    ||||||     | ||| | || ||| || ||    ||
Db       8734 GAGTCGGCGGTGTTCGCGGAGCGGATTGCCGAGTGTGCGGCGGCACTCGCCGAGTTCACC 8793

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||  |||| ||  | |||||  | |  | ||| |  ||  || |||||||| |
Db       8794 GACTGGTCGCTGGTCGATGTGCTGCGGGGTGTGGTGGGTGCGCCGTCGTTGGAGCGGGTC 8853

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              |||||||| ||||| | |  ||||   || |||||||| ||||||||  ||||| |||||
Db       8854 GATGTGGTGCAGCCGGCGTCGTTCGCGGTGATGGTGTCGTTGGCTGCGTTGTGGGGTTCC 8913

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
                ||| |    || || ||||| | || ||||| ||||| ||||||||  ||||||
Db       8914 CGTGGTGTGTTGCCGGATGCGGTGGTGGGGCATTCGCAGGGTGAGATCGCTGCCGCGGTG 8973

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||||  || ||||||    ||  ||||| |||| |   || |||||||| || | |||
Db       8974 GTGTCGGGTGCGCTGTCGTTGCGGGACGGGGCGCGGGTGGTGGCGCTGCGGAGTCAGGCC 9033

Qy        361 CTGG---CCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAG 417
              | |      ||  ||  ||| |||||||| |||   || ||  ||||| | |   | | |
Db       9034 ATTGGTCGTGCGTTGGCGGGGCGGGGCGGGATGATGTCCGTCGCGCTGTCGGTGGACGTG 9093

Qy        418 GTGGAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGC 477
              | ||| | ||   | | ||   ||  | ||||||| ||| |||||| || |||||||||
Db       9094 CTCGAACCGCGGTTGGTCGA---GTTCGAGGGGCGGGTGTCGGTGGCCGCCGTCAACGGC 9150

Qy        478 CCCCGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGT 537
              || ||||||  ||  ||| | || ||  |||||||| |||||| | |||  ||
Db       9151 CCGCGCTCCGTCGTGGTCGCCGGCGAGCCCGAGGCGCTGGACGCGCTGCACGCCCGGCTG 9210

Qy        538 GCCGGCACCGGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCAT 597
              ||| |  ||   | |||||||| |||||| |||| ||||| ||||||||||  | |||
Db       9211 ACCGCCGACGACATCCGGGCCCGCCGGATCGCGGTGGACTACGCCTCGCACTCGCACCAG 9270

Qy        598 GTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCC 657
              ||  ||  |||||  |||||  |||||||| |||||| |||  |     ||||   ||||
Db       9271 GTCGAGGACCTGCACGAGGAACTGCTGGAGGTGCTGGCGGAGCTGGCGCCGCGCACGTCG 9330

Qy        658 GGCGTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCC 717
              |  |||||||||||||| || |||   |||  ||||||||||||  |     ||||||||
Db       9331 GAGGTGCCGTTCTTCTCGACCGTGACCGGCGACTGGCTGGACACCGCGCGGATGGACGCC 9390
```

```
Qy        718 GCCTACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTG 777
              | |||||||| |||||||||||| |    | ||| || |||    || || ||    || ||||
Db       9391 GGCTACTGGTTCCGCAACCTGCGCGGACGGGTGCGGTTCGCGGACGCGGTGGCGGACCTG 9450

Qy        778 GCGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
                ||  |  |   ||||||  |||||||| |||||| | |||||    ||       ||
Db       9451 CTGGCGGCGGAGTACCGCGCGTTCGTCGAGGTCAGCTCGCACCCGGTGCTGACGATGGCG 9510

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897
              ||   |||   | ||| ||  |||  |     | |  ||||| ||||| ||| |||||
Db       9511 GTCTTGGACCTGATCGAGGAGGCCGGGGTCACGGCCGTCGCGACCGGCACCCTGCGCCGT 9570

Qy        898 GGCGACAACGACACCCGCCGCTTC 921
              | | |    | | |  ||||||||
Db       9571 GACCAGGGTGGCGCGGGCCGCTTC 9594
```

```
RESULT 8
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag=  a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag=  b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag=  c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag=  d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag=  e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag=  f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag=  g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag=  h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
```

```
  Query Match           43.7%;  Score 402.4;  DB 22;  Length 125401;
  Best Local Similarity   66.2%;  Pred. No. 2.2e-59;
  Matches  612;  Conservative    0;  Mismatches  306;  Indels    6;  Gaps    2;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              || |||||||| |||| |||||| |||||||| |||||||| |   |  || ||||
Db      68935 GTGTTCGTGTTCCCCGGTCAGGGTTCGCAGTGGGTGGGGATGGGGGCCCAACTCCTCGAT 68994

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
                ||  |||||||||    ||||||    | ||| | || ||| || ||   ||
Db      68995 GAGTCGGCGGTGTTCGCGGAGCGGATTGCCGAGTGTGCGGCGGCACTCGCCGAGTTCACC 69054

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||  |||| ||  | |||||  | |  | ||| |    ||  || |||||||| |
Db      69055 GACTGGTCGCTGGTCGATGTGCTGCGGGGTGTGGTGGGTGCGCCGTCGTTGGAGCGGGTC 69114

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              |||||||| ||||| | |  ||||    || |||||||| ||||||||  ||||| |||||
Db      69115 GATGTGGTGCAGCCGGCGTCGTTCGCGGTGATGGTGTCGTTGGCTGCGTTGTGGGGTTCC 69174

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              |||  |    || || |||||  | || |||||  |||||  |||||||| ||||||
Db      69175 CGTGGTGTGTTGCCGGATGCGGTGGTGGGGCATTCGCAGGGTGAGATCGCTGCCGCGGTG 69234

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||||  || ||||||    ||  ||||| ||||  |   || |||||||| || | |||
Db      69235 GTGTCGGGTGCGCTGTCGTTGCGGGACGGGGCGCGGGTGGTGGCGCTGCGGAGTCAGGCC 69294

Qy        361 CTGG---CCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAG 417
              | |      ||  ||  ||| |||||||| |||   || ||  ||||| | |   | | |
Db      69295 ATTGGTCGTGCGTTGGCGGGGCGGGGCGGGATGATGTCCGTCGCGCTGTCGGTGGACGTG 69354

Qy        418 GTGGAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGC 477
              | || | ||   | |  ||   ||  | ||||||| ||| |||||| || |||||||||
Db      69355 CTCGAACCGCGGTTGGTCGA---GTTCGAGGGGCGGGTGTCGGTGGCCGCCGTCAACGGC 69411

Qy        478 CCCCGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGT 537
              || |||||  ||  ||| | || ||  |||||||| |||||| | |||  ||
Db      69412 CCGCGCTCCGTCGTGGTCGCCGGCGAGCCCGAGGCGCTGGACGCGCTGCACGCCCGGCTG 69471

Qy        538 GCCGGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCAT 597
              ||| |  ||   | |||||||| |||||| |||| |||| ||||||||||   | |||
Db      69472 ACCGCCGACGACATCCGGGCCCGCCGGATCGCGGTGGACTACGCCTCGCACTCGCACCAG 69531
```

```
Qy        598 GTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCC 657
              ||  ||  |||||  |||||  |||||||| ||||||  |||  |    ||||   ||||
Db      69532 GTCGAGGACCTGCACGAGGAACTGCTGGAGGTGCTGGCGGAGCTGGCGCCGCGCACGTCG 69591

Qy        658 GGCGTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCC 717
              |  |||||||||||||| || |||   |||  |||||||||||| |      ||||||||
Db      69592 GAGGTGCCGTTCTTCTCGACCGTGACCGGCGACTGGCTGGACACCGCGCGGATGGACGCC 69651

Qy        718 GCCTACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTG 777
              | ||||||| ||||||||||| |    | ||| || |||   || || ||   || ||||
Db      69652 GGCTACTGGTTCCGCAACCTGCGCGGACGGGTGCGGTTCGCGGACGCGGTGGCGGACCTG 69711

Qy        778 GCGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
              ||  |  |   ||||||  |||||||| ||||||  | |||||     ||        ||
Db      69712 CTGGCGGCGGAGTACCGCGCGTTCGTCGAGGTCAGCTCGCACCCGGTGCTGACGATGGCG 69771

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897
              ||  |||   | ||| ||  |||  |     | |  ||||| ||||| ||| |||||
Db      69772 GTCTTGGACCTGATCGAGGAGGCCGGGGTCACGGCCGTCGCGACCGGCACCCTGCGCCGT 69831

Qy        898 GGCGACAACGACACCCGCCGCTTC 921
              | | |   | | |  |||||||
Db      69832 GACCAGGGTGGCGCGGGCCGCTTC 69855


RESULT 9
AAD17185
ID   AAD17185 standard; DNA; 27541 BP.
XX
AC   AAD17185;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys2; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (454..1191)
FT                   /*tag= a
FT                   /product= "NysF protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (1275..3092)
FT                   /*tag= b
FT                   /product= "NysG protein"
FT   CDS             complement (3070..4824)
FT                   /*tag= c
FT                   /note= "CDS does not include start codon"
FT                   /product= "NysH protein"
FT   CDS             5122..6156
FT                   /*tag= d
FT                   /product= "NysD3 protein"
FT   CDS             6338..27541
FT                   /*tag= e
FT                   /product= "NysI partial protein"
FT                   /note= "CDS does not include stop codon"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 151-166; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

  Query Match           43.5%;  Score 400.8;  DB 22;  Length 27541;
  Best Local Similarity   66.1%;  Pred. No. 4.4e-59;
  Matches  611;  Conservative    0;  Mismatches  307;  Indels    6;  Gaps    2;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              || |||||||| ||||| ||||||  ||||||||  ||||||||     |  || || |
Db      18020 GTGTTCGTGTTCCCCGGTCAGGGTTCGCAGTGGGTGGGGATGGGGTCCCAACTCCTTGAT 18079

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
                ||  ||||||||||   ||||||    | ||| | || ||| || ||   ||
Db      18080 GAGTCGGCGGTGTTCGCGGAGCGGATTGCCGAGTGTGCGGCGGCACTCGCCGAGTTCACC 18139

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||||  |||| ||  | |||||  | |  | ||| |   ||  || ||||||| |
Db      18140 GACTGGTCGCTGGTCGATGTGCTGCGGGGTGTGGTGGGTGCGCCGTCGTTGGAGCGGGTC 18199

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              |||||||| ||||| | |  |||   || |||||||| ||||||||  ||||||||||
Db      18200 GATGTGGTGCAGCCGGCGTCGTTCGCGGTGATGGTGTCGTTGGCTGCGTTGTGGCGTTCC 18259

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
                ||| |    || || |||||  | || ||||| ||||| |||||||| || |||
Db      18260 CGTGGTGTGTTGCCGGATGCGGTGGTGGGGCATTCGCAGGGTGAGATCGCTGCTGCGGTG 18319

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||||  || ||||||     ||  ||||| ||||  |    || |||||||| || | |||
Db      18320 GTGTCGGGTGCGCTGTCGTTGCGGGACGGGGCGCGGGTGGTGGCGCTGCGGAGTCAGGCC 18379

Qy        361 CTGG---CCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAG 417
              | |      ||  ||  ||| ||||| || |||   || ||  ||||| | |   | | |
Db      18380 ATTGGTCGTGCGTTGGCGGGGCGGGGAGGGATGATGTCCGTCGCGCTGTCGGTGGACGTG 18439

Qy        418 GTGGAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGC 477
              | || | ||   | |  ||   ||  | ||||||| ||| |||||| || |||||||||
Db      18440 CTCGAACCGCGGTTGGTCGA---GTTCGAGGGGCGGGTGTCGGTGGCCGCCGTCAACGGC 18496

Qy        478 CCCCGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGT 537
              || ||||||  ||  ||| | || ||  |||||||| |||||| | |||  ||
Db      18497 CCGCGCTCCGTCGTGGTCGCCGGCGAGCCCGAGGCGCTGGACGCGCTGCACGCCCGGCTG 18556

Qy        538 GCCGGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCAT 597
              ||| |  ||   | |||||||| |||||| |||| |||||  |||||||||  | |||
Db      18557 ACCGCCGACGACATCCGGGCCCGCCGGATCGCGGTGGACTACGCCTCGCACTCGCACCAG 18616

Qy        598 GTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCC 657
              ||  ||  |||||  |||||  |||||||| ||||||  |||  |    ||||   ||||
```

```
Db      18617 GTCGAGGACCTGCACGAGGAACTGCTGGAGGTGCTGGCGGAGCTGGCGCCGCGCACGTCG 18676

Qy        658 GGCGTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCC 717
              |  |||||||||||||| || |||    |||  ||||||||||||| |      ||||||||
Db      18677 GAGGTGCCGTTCTTCTCGACCGTGACCGGCGACTGGCTGGACACCGCGCGGATGGACGCC 18736

Qy        718 GCCTACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTG 777
              | |||||||| |||||||||||  |    | ||| || |||   || || ||   || ||||
Db      18737 GGCTACTGGTTCCGCAACCTGCGCGGACGGGTGCGGTTCGCGGACGCGGTGGCGGACCTG 18796

Qy        778 GCGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
                ||  |  |   ||||||  |||||||| |||||| | |||||     ||        ||
Db      18797 CTGGCGGCGGAGTACCGCGCATTCGTCGAGGTCAGCTCGCACCCGGTGCTGTCGATGGCG 18856

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897
              |  | ||  | | ||| ||  ||| ||      | |  |||   ||||| ||| ||||||
Db      18857 GTGCAGGAGGCGATCGACGAGGCCGGCGTGCCGGCCGTCGCCGCCGGCACCCTGCGCCGC 18916

Qy        898 GGCGACAACGACACCCGCCGCTTC 921
              | | |   || ||||  |||||||
Db      18917 GACCAGGGCGGCACCGACCGCTTC 18940


RESULT 10
AAQ46806
ID   AAQ46806 standard; DNA; 29879 BP.
XX
AC   AAQ46806;
XX
DT   22-DEC-1993  (first entry)
XX
DE   eryA region of S. erythraea chromosome.
XX
KW   Saccarapolyspora erythraea; eryA; biosynthesis; polyketide; module;
KW   erythromycin; condensation; elongation; acyl chain growth;
KW   gene replacement; ss.
XX
OS   Saccarapolyspora erythraea.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..10199
FT                   /*tag=  a
FT                   /note= "ORF 1"
FT   CDS             10218..20921
FT                   /*tag=  b
FT                   /note= "ORF 2"
FT   CDS             20922..29879
FT                   /*tag=  c
FT                   /note= "ORF 3"
XX
PN   WO9313663-A.
XX
PD   22-JUL-1993.
XX
PF   17-JAN-1992;   92WO-US00427.
XX
PR   17-JAN-1992;   92WO-US00427.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Donadio S;  Katz L,  McAlpine JB;
XX
DR   WPI; 1993-242804/30.
DR   P-PSDB; AAR44430-32.
XX
PT   Biosynthesis of specific polyketide analogues esp. erythromycin
PT   cpds. - by introducing altered biosynthetic gene-contg. DNA into
PT   microorganisms
XX
PS   Claim 27; Fig 2; 133pp; English.
XX
CC   This sequence represents a fragment of the Saccarapolyspora erythraea
CC   genome, designated eryA.  The polypeptides encoded by this region
CC   are involved in the biosynthesis of the polyketide segment of
CC   erythromycin.  eryA is organised in modules and each module takes care
CC   of one condensation step.  The precise succession of elongation steps
CC   is dictated by the genetic order of the modules.  This fragment may be
CC   specifically altered such that novel polyketide molecules of desired
CC   structure are produced.  Three types of alteration may be produced;
CC   those inactivating a single function in a module which does not arrest
CC   acyl chain growth; those inactivating a single function in a module
CC   which does affect chain growth; and those affecting an entire module.
CC   The mutations may be introduced by gene replacement.
XX
SQ   Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 other;

  Query Match             41.8%;  Score 385.4;  DB 14;  Length 29879;
  Best Local Similarity   65.6%;  Pred. No. 1.7e-56;
  Matches  579;  Conservative    0;  Mismatches  301;  Indels    3;  Gaps    1;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              |||||| |||| || || |||||  ||| ||||||| |||||||     |||||  |
Db      11895 GTCTTCCTGTTCCCGGGGCAGGGTTCGCAATGGGCCGGCATGGGTGCCGAGCTGCTGTCG 11954

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              || || |||||||||||| ||  ||| | ||| ||||| ||| |  |||| |||  |  |
Db      11955 TCGTCGCCGGTGTTCGCCGGGAAGATCCGGGCCTGCGACGAGTCGATGGCCCCGATGCAG 12014

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              ||||||   ||    ||| | ||||| | ||   ||||||   |||    ||| |||| |
Db      12015 GACTGGAAGGTCTCCGACGTGCTGCGTCAGGCGCCGGGGGCGCCGGGCCTGGACCGGGTC 12074

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              || ||||| ||||| ||| |||||    || ||||||||  |||| |  |||||||| ||
Db      12075 GACGTGGTGCAGCCGGTGTTGTTCGCGGTGATGGTGTCGCTGGCGGAGCTGTGGCGCTCG 12134

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              |||||  | || ||||  |||||| | || || || ||||||||||||| ||||||||
Db      12135 TACGGCGTGGAGCCCGCGGCGGTCGTGGGGCACTCGCAGGGCGAGATCGCCGCCGCGCAC 12194

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||    |||||||| |   || ||||||||||||||   || | |  ||||||||   |
Db      12195 GTCGCCGGGGCGCTCACGTTGGAGGACGCGGCGAAGCTCGTCGTGGGCCGCAGCCGCCTG 12254

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              ||    |  |   ||  |||||||||||| | ||  |||||   |    | ||||
Db      12255 ATGCGGTCGCTCTCCGGGGAGGGCGGCATGGCCGCCGTCGCGCTGGGCGAGGCCGCGGTG 12314

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
                 ||| | |  |   || |||| ||  ||| | | |||||| ||||||||||| |||
Db      12315 CGCGAGCGCCTGCG---GCCGTGGCAGGACCGGCTCTCGGTGGCCGCGGTCAACGGTCCC 12371

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              || ||   ||  ||||| || ||  |||  ||| ||   | | |    | | |||| ||
Db      12372 CGGTCGGTCGTGGTCTCCGGCGAGCCCGGCGCGCTGCGGGCGTTTTCCGAGGACTGCGCG 12431

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              | |   ||  | || | |||    |||   || ||||| ||||||||||  || ||  |
Db      12432 GCCGAGGGCATCCGCGTCCGCGACATCGACGTGGACTACGCCTCGCACTCGCCGCAGATC 12491

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              |||   | || |||||  | || ||   |   || ||||||   ||||  ||| | |
Db      12492 GAGCGGGTCCGCGAGGAACTCCTCGAAACGACCGGCGACATCGCGCCGCGCCCGGCGCGG 12551

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              ||| |||||  |||||| ||||||       |  |||||  |||    ||||| |||
Db      12552 GTGACGTTCCACTCCACTGTGGAGTCGCGGTCTATGGACGGCACCGAGCTGGATGCCCGG 12611

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              |||||||||||||||||| | || |||| || |||  |||  |||||   |   ||||||
Db      12612 TACTGGTACCGCAACCTGCGCGAGACGGTGCGCTTCGCCGACGCCGTGACGCGGCTGGCG 12671

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              ||    ||| ||  ||  ||| |||| |||||||| || ||   | | ||||  ||| ||
Db      12672 GAGTCGGGATACGACGCGTTCATCGAGGTCAGCCCGCATCCGGTCGTGGTCCAGGCCGTC 12731
```

```
Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCG 883
              || ||  |   ||||||  | | ||  | ||    |||| |||
Db      12732 GAGGAGGCGGTCGAAGAGGCTGACGGTGCCGAAGACGCGGTCG 12774
```

RESULT 11
AAA09469
ID AAA09469 standard; DNA; 50937 BP.
XX
AC AAA09469;
XX
DT 29-AUG-2000 (first entry)
XX
DE Streptococcus oleandomycin gene cluster.
XX
KW oleandomycin; oleandolide; polyketide synthase; oleAI; oleAII; oleAIII;
KW PKS; type I; 8,8a-deoxyoleandolide synthase; modular; ketosynthase;
KW acyl-transferase; acyl carrier protein; inactivated; polyketide;
KW macrolactone; antibiotic; motilide; erythromycin; ss.
XX
OS Streptococcus antibioticus.
XX

FH	Key	Location/Qualifiers
FT	CDS	152..1426
FT		/*tag= a
FT		/label= oleI
FT	CDS	complement (1528..2637)
FT		/*tag= b
FT		/label= oleN2
FT	CDS	complement (2658..4967)
FT		/*tag= c
FT		/label= oleR
FT	CDS	5772..18224
FT		/*tag= d
FT		/label= ORF1
FT		/product= 8,8a-deoxyoleandolide_synthase_1
FT	CDS	18267..29717
FT		/*tag= e
FT		/label= ORF2
FT		/product= 8,8a-deoxyoleandolide_synthase_2
FT	CDS	29787..40346
FT		/*tag= f
FT		/label= ORF3
FT		/product= 8,8a-deoxyoleandolide_synthase_3
FT	CDS	40625..41830
FT		/*tag= g
FT		/label= oleP1
FT	CDS	41878..43158
FT		/*tag= h
FT		/label= oleG1
FT	CDS	43163..44443
FT		/*tag= i
FT		/label= oleG2
FT	CDS	44433..45173
FT		/*tag= j
FT		/label= oleM1
FT	CDS	45251..46411
FT		/*tag= k
FT		/label= oleY
FT	CDS	46491..47714
FT		/*tag= l
FT		/label= oleP
FT	CDS	complement (47808..49517)
FT		/*tag= m
FT		/label= oleB

XX
PN WO200026349-A2.
XX
PD 11-MAY-2000.
XX
PF 22-OCT-1999; 99WO-US24478.
XX
PR 29-OCT-1998; 98US-0106100.
PR 16-FEB-1999; 99US-0120254.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Betlach MC, Shah SK, McDaniel R, Tang L;
XX
DR WPI; 2000-365602/31.
DR P-PSDB; AAY92707, AAY92708, AAY92709.
XX
PT Recombinant DNA compound encoding oleandolide polyketide synthase for
PT synthesizing polyketides comprising a coding sequence for a domain of a
PT loading module or any one of extender modules
XX
PS Disclosure; Page 14-26; 86pp; English.
XX
CC This is part of the Streptococcus antibioticus oleandomycin gene cluster.
CC The oleandolide polyketide synthase (PKS), also known as
CC 8,8a-deoxyoleandolide synthase, is encoded by three open reading frames
CC (ORF), designated oleAI, oleAII and oleAIII. The PKS is a type I
CC "modular" enzyme, where each ORF encodes 2 extender modules and
CC the first ORF also encodes the loading module. Each module is composed
CC of at least a ketosynthase (KS), acyl-transferase (AT) and an
CC acyl carrier protein (ACP) domain. The oleandolide PKS loading module
CC contains an inactivated KS, called KS-Q, where Q is the abbreviation for
CC glutamine, present instead of the active site cysteine required for
CC activity. The large multifunctional PKS enzymes catalyze the biosynthesis
CC of polyketide macrolactones through multistep pathways involving
CC decarboxylative condensations between acylthioesters followed by cycles
CC of varying beta-carbon processing activities. The macrolide product of
CC the PKS, 8,8a-deoxyoleandolide, is further modified by epoxidation and
CC glycosylation to yield oleandomycin, an antibacterial polyketide. The
CC invention concerns an isolated recombinant DNA compound, comprising a
CC coding sequence for a domain of loading module or any one of extender
CC modules 1-4 or 1-6, including an oleandolide PKS operably linked to a
CC promoter. Also discussed are recombinant oleandolide PKS in which the
CC module 1 KS domain is inactivated by deletion or other mutation. In
CC particular, the inactivation is mediated by a change in the KS domain
CC that renders it incapable of binding substrate (the KS1-o mutation),
CC rendered by mutation in the codon for the active site cysteine. The
CC oleandolide PKS is useful for synthesizing polyketides, which are useful
CC as antibiotics and motilides. Heterologous expression of oleandolide PKS
CC in host cells such as Streptomyces coelicolor and S. lividans is also
CC made possible. Unmodified oleandolide compounds can be provided to
CC cultures of Saccharopolyspora erythraea and converted to the
CC corresponding derivatives of erythromycins A-D.
XX
SQ Sequence 50937 BP; 6672 A; 16253 C; 19272 G; 8740 T; 0 other;

Query Match 40.6%; Score 373.8; DB 21; Length 50937;
Best Local Similarity 64.4%; Pred. No. 1.5e-54;
Matches 593; Conservative 0; Mismatches 322; Indels 6; Gaps 2;

```
Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              || ||||| ||||| || ||||||  ||||||||  |||||||  |    |  |||| |
Db      24645 GTGTTCGTCTTTCCTGGTCAGGGTTCGCAGTGGGTGGGGATGGCGGCTGGGTTGCTGGAT 24704

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
                | | ||||||||||||||   |  | ||  || ||| |  | ||  ||||  |||  |  |
Db      24705 GCGTGTCCGGTGTTCGCGGAGGCGGTGGCGGAGTGTGCCGCGGTGCTGGATCCGGTGACG 24764

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              |  |||||  |||| ||  |  |||||    |  |  || || || ||    | |||||
Db      24765 GGTTGGTCGCTGGTCGAGGTGTTGCGC---GGTGGTGGTGAGGCTGTTCTTGGGCGGGTT 24821

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              |||||||| |||||  |  | |||      || ||||||||  ||||        |||||
Db      24822 GATGTGGTGCAGCCGGCGTTGTGGGCGGTGATGGTGTCACTGGCCCGGACCTGGCGGTAT 24881

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              ||||||  | ||  ||  |  |||||  | ||  |||||  |||||  ||||||||  ||  ||    |
Db      24882 TACGGTGTGGAGCCTGCTGCGGTTGTGGGGCATTCGCAGGGTGAGATCGCTGCGGCTTGT 24941

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
```

```
Db      24942 GTGGCTGGGGGGTTGAGTCTGGCCGATGGTGCGCGGGTGGTGGTGTTGCGGAGCCGGGCG 25001

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420

Db      25002 ATCGCCCGGATCGCTGGTGGGGGCGGCATGGTCTCGGTCGGTCTTTCAGC---TGAGCGT 25058

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480

Db      25059 GTCCGCACCATGCTCGACACCTACGGTGGCCGGGTTTCGGTCGCGGCGGTCAATGGCCCG 25118

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540

Db      25119 TCCTCGACCGTCGTGTCCGGTGACGTCCAGGCCCTGGATGAGTTGTTGGCCGGTTGTGAG 25178

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600

Db      25179 CGGGAGGGTGTCCGGGCTCGTCGTGTCCCGGTGGACTATGCCTCCCACTCCGCGCAGATG 25238

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660

Db      25239 GACCAGTTACGCGATGAGCTGCTGGAAGCGCTGGCGGACATCACTCCGCAACATTCCAGT 25298

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720

Db      25299 GTTCCGTTCTTCTCGACGGTGACGGCGGACTGGCTGGACACGACCGCTCTGGATGCGGGG 25358

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780

Db      25359 TACTGGTTCACGAATCTGCGGGAGACGGTCCGGTTCCAGGAAGCCGTCGAAGGGCTCGTG 25418

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840

Db      25419 GCTCAGGGGATGGGCGCGTTCGTCGAGTGCAGCCCGCACCCCGTCCTCGTCCCCGGTATC 25478

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900

Db      25479 GAGCAGACCCTCGACGCCCTCGACCAGAACGCCGCCGTACTCGGCTCCCTGCGGCGTGAC 25538

Qy        901 GACAACGACACCCGCCGCTTC 921

Db      25539 GAAGGCGGCCTGGACCGACTC 25559


RESULT 12
AAV99255
ID   AAV99255 standard; DNA; 1010 BP.
XX
AC   AAV99255;
XX
DT   09-MAR-1999  (first entry)
XX
DE   DNA encoding an active acyltransferase domain (NidAT5) from a PKS.
XX
KW   Enzymatically active acyltransferase domain; polyketide synthase; PKS;
KW   polyketide synthesis; antimicrobial; anticancer; antifungal;
KW   immunosuppressant; anthelmintic; NidAT5; ethyl acyltransferase domain;
KW   module 5; PKS cluster; ds.
XX
OS   Streptomyces caelestis.
XX
PN   WO9851695-A2.
XX
PD   19-NOV-1998.
XX
PF   13-MAY-1998;   98WO-US09518.
XX
PR   16-MAY-1997;   97US-0858003.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Kakavas SJ,  Katz L,  Pereda-Lopez A,  Ruan X,  Stassi DL;
PI   Summers RG;
XX
DR   WPI; 1999-070114/06.
DR   P-PSDB; AAW87714.
XX
PT   New polyketides produced by microorganism having altered polyketide
PT   synthase genes - especially chimeras containing a heterologous
PT   acyltransferase domain, useful as e.g. antimicrobials, anticancer
PT   agents
XX
PS   Claim 24; Fig 21; 137pp; English.
XX
CC   The present sequence encodes an enzymatically active acyltransferase
CC   domain, NidAT5, from a polyketide synthase (PKS). The NidAT5 protein
CC   is derived from the ethyl acyltransferase domain from module 5 of the
CC   PKS cluster of Streptomyces caelestis NRRL-2821. The sequence
CC   can be used to replace acyltransferase domains of other PKS enzymes,
CC   which are then used to generate novel polyketides by de novo biosynthesis
CC   rather than by chemical modification. Polyketides of the invention have
CC   the methyl groups on the macrolactone ring substituted with -H, -Et
CC   and/or -OH, and the ethyl side chain substituted with hydroxymethyl or
CC   dihydroxycyclohexylmethyl side chains. Polyketides of the invention are
CC   antimicrobial, anticancer, antifungal, immunosuppressant and/or
CC   anthelmintic agents, or their intermediates.
XX
SQ   Sequence 1010 BP; 104 A; 353 C; 402 G; 151 T; 0 other;

  Query Match             39.4%;  Score 362.8;  DB 20;  Length 1010;
  Best Local Similarity   63.7%;  Pred. No. 1.3e-52;
  Matches  585;  Conservative    0;  Mismatches  327;  Indels    6;  Gaps    2;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60

Db         13 GTGTTCGTGTTCCCCGGCCAGGGCTCGCAGTGGGCCGGAATGGCCGAGGGGCTGCTGGAG 72

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120

Db         73 CGGTCCGGCGCGTTCCGGAGTGCGGCCGACTCGTGCGACGCCGCGCTGCGGCCGTACCTC 132

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180

Db        133 GGCTGGTCGGTGCTGAGCGTGCTGCGCGGGGAACCGGACGCGCCCTCGCTCGACCGGGTC 192

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240

Db        193 GACGTCGTGCAGCCGGTGCTGTTCACGATGATGGTCTCGCTCGCGGCGGTCTGGCGTGCG 252

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300

Db        253 CTGGGGGTGGAACCGGCGGCGGTCGTCGGGCACTCGCAGGGTGAGATCGCCGCTGCCCAT 312

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360

Db        313 GTCGCCGGTGCGCTGTCGCTGGACGACTCGGCCCGGATCGTCGCCCTGCGCAGTCGGGCG 372

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420

Db        373 TGGCTCGGACTGGCGGGCAAGGGCGGCATGGTGGCGGTGCCGATGCCGGCGGAGGAGCTG 432

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480

Db        433 CGGCCGCGGCT---GGTGACGTGGGGGGACCGTCTGGCCGTCGCCGCCGTCAACAGCCCC 489

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540

Db        490 GGTTCCTGCGCCGTCGCAGGCGACCCGGAGGCGCTGGCCGAACTGGTGGCGCTGCTGACC 549

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCC---GGTCGACTATGCCTCGCACTGCCCCCAT 597

Db        550 GGTGAGGGGGTGCACGCCCGGCCGATCCCCGGCGTCGACACGGCGGGCCACTCGCCGCAG 609

Qy        598 GTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCC 657

Db        610 GTGGACGCGTTGCGGGCTCATCTGCTGGAGGTGCTGGCCCCGGTCGCCCCCCGACCGGCC 669
```

```
Qy        658 GGCGTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCC 717
              | | | ||||||| ||| ||||||   |||     |||||||  |||    ||||||||
Db        670 GACATCCCGTTCTACTCGACGGTGACCGGCGGGCTGCTGGACGGCACCGAGCTGGACGCG 729

Qy        718 GCCTACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTG 777
              | ||||||||||||||| ||| | |||| |||   |||       |||   | ||| |||
Db        730 ACGTACTGGTACCGCAACATGCGCGAGCCCGTCGAGTTCGAGCGGGCCACACGGGCGCTG 789

Qy        778 GCGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
                | ||||| |||  |||||||  | ||    |||||  || ||||  || | |   ||
Db        790 ATCGCCGACGGGCACGACGTCTTCCTGGAGACGAGCCCGCATCCCATGCTGGCCGTGGCG 849

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897
              | ||  | ||   |   ||| ||| |   ||||   | | | |||| | ||| ||||||
Db        850 CTGGAGCAGACGGTCACCGACGCCGGCACCGACGCGGCGGTGCTCGGGACCCTGCGCCGC 909

Qy        898 GGCGACAACGACACCCGC 915
              || ||  ||   | |||
Db        910 CGCCACGGCGGTCCTCGC 927
```

RESULT 13
AAA72521
ID AAA72521 standard; DNA; 1010 BP.
XX
AC AAA72521;
XX
DT 20-NOV-2000 (first entry)
XX
DE DNA sequence of the ethyl AT module of the niddamycin PKS cluster.
XX
KW Erythromycin derivative; polyketide synthase; PKS; acetyltransferase;
KW antibiotic; antifungal; anticancer; immunosuppressant; antihelminithic;
KW ethyl AT; niddamycin; ds.
XX
OS Streptomyces caelestis.
XX
PN US6063561-A.
XX
PD 16-MAY-2000.
XX
PF 13-MAY-1998; 98US-0078166.
XX
PR 17-JAN-1991; 91US-0642734.
PR 16-MAY-1997; 97US-0858003.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Katz L, Stassi DL, Ruan X, Pereda-Lopez A, Kakavas SJ, Summers RG;
XX
DR WPI; 2000-386760/33.
DR P-PSDB; AAB12926.
XX
PT New erythromycin derivatives produced by microorganisms with altered
PT polyketide synthase genes -
XX
PS Example 20; Fig 21; 94pp; English.
XX
CC This invention relates to novel erythromycin derivatives, in which the
CC methyl groups on the macrolactone ring have been substituted with H, Et
CC and/or OH and the ethyl side chain has been substituted with
CC hydroxymethyl or dihydroxycyclohexylmethyl side chains. Included in the
CC invention are vectors containing polynucleotide sequences which encode an
CC enzymatically active acetyltransferase domain from Streptomyces. Also
CC included are methods for altering the substrate specificity of a
CC polyketide synthase in a microorganism using the vectors of the
CC invention. Polyketides are are a large class of natural products
CC including antibiotics, antifungal, anticancer and immunosuppressant
CC compounds. Polyketide synthase (PKS) is a complex of enzyme activities
CC responsible for the biosynthesis of polyketides. The erythromycin
CC derivatives exhibit antibiotic, antifungal, anticancer, antihelminithic
CC and immunosuppressant activity.
CC The present sequence represents the DNA sequence encoding the ethyl
CC acetyltransferase (AT) module of the niddamycin PKS cluster isolated from
CC Streptomyces caelestis. The sequence is used in the production of a
CC vector of the invention.
XX
SQ Sequence 1010 BP; 104 A; 353 C; 402 G; 151 T; 0 other;

Query Match 39.4%; Score 362.8; DB 21; Length 1010;
Best Local Similarity 63.7%; Pred. No. 1.3e-52;
Matches 585; Conservative 0; Mismatches 327; Indels 6; Gaps 2;

```
Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              || |||||||| ||||| |||||   |||||||||||| ||||  | | | ||||| |
Db         13 GTGTTCGTGTTCCCCGGCCAGGGCTCGCAGTGGGCCGGAATGGCCGAGGGGCTGCTGGAG 72

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
                ||     | ||||     |   |     | ||||||| |  || |||  |||||    |
Db         73 CGGTCCGGCGCGTTCCGGAGTGCGGCCGACTCGTGCGACGCCGCGCTGCGGCCGTACCTC 132

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              | |||||| ||| ||  | | |||||| ||||  |||  |    |    |   || |||| |
Db        133 GGCTGGTCGGTGCTGAGCGTGCTGCGCGGGGAACCGGACGCGCCCTCGCTCGACCGGGTC 192

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              || || || ||||| ||||||||||    | ||||| ||  | ||  ||  | |||||| |
Db        193 GACGTCGTGCAGCCGGTGCTGTTCACGATGATGGTCTCGCTCGCGGCGGTCTGGCGTGCG 252

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
                || | ||||| |  |||||| | || || || ||||| ||||||||  || || |||
Db        253 CTGGGGGTGGAACCGGCGGCGGTCGTCGGGCACTCGCAGGGTGAGATCGCCGCTGCCCAT 312

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||    || ||||||   ||| | ||| ||||   ||  || || |||||||| ||||||
Db        313 GTCGCCGGTGCGCTGTCGCTGGACGACTCGGCCCGGATCGTCGCCCTGCGCAGTCGGGCG 372

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              |  ||   ||  |||| ||||||||||   | |||||| |||| ||  ||||| ||
Db        373 TGGCTCGGACTGGCGGGCAAGGGCGGCATGGTGGCGGTGCCGATGCCGGCGGAGGAGCTG 432

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              || ||   |  | |  ||||| ||  ||  ||   || || || |||||| |||||
Db        433 CGGCCGCGGCT---GGTGACGTGGGGGGACCGTCTGGCCGTCGCCGCCGTCAACAGCCCC 489

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              | |||  ||||||| | || ||  | |||||| ||| |||  || |||||        ||
Db        490 GGTTCCTGCGCCGTCGCAGGCGACCCGGAGGCGCTGGCCGAACTGGTGGCGCTGCTGACC 549

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCC---GGTCGACTATGCCTCGCACTGCCCCCAT 597
              ||    ||||||| ||||||| |||||    ||||||   ||    ||||  || ||
Db        550 GGTGAGGGGGTGCACGCCCGGCCGATCCCCGGCGTCGACACGGCGGGCCACTCGCCGCAG 609

Qy        598 GTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCC 657
              ||| |  |  |||||   | |||||||| |||||        ||  ||| |  ||| ||
Db        610 GTGGACGCGTTGCGGGCTCATCTGCTGGAGGTGCTGGCCCCGGTCGCCCCCCGACCGGCC 669

Qy        658 GGCGTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCC 717
              | | | ||||||| ||| ||||||   |||     |||||||  |||    ||||||||
Db        670 GACATCCCGTTCTACTCGACGGTGACCGGCGGGCTGCTGGACGGCACCGAGCTGGACGCG 729

Qy        718 GCCTACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTG 777
              | ||||||||||||||| ||| | |||| |||   |||       |||   | ||| |||
Db        730 ACGTACTGGTACCGCAACATGCGCGAGCCCGTCGAGTTCGAGCGGGCCACACGGGCGCTG 789

Qy        778 GCGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCC 837
                | ||||| |||  |||||||  | ||    |||||  || ||||  || | |   ||
Db        790 ATCGCCGACGGGCACGACGTCTTCCTGGAGACGAGCCCGCATCCCATGCTGGCCGTGGCG 849

Qy        838 ATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGC 897
              | ||  | ||   |   ||| ||| |   ||||   | | | |||| | ||| ||||||
Db        850 CTGGAGCAGACGGTCACCGACGCCGGCACCGACGCGGCGGTGCTCGGGACCCTGCGCCGC 909

Qy        898 GGCGACAACGACACCCGC 915
              || ||  ||   | |||
```

```
Db        910 CGCCACGGCGGTCCTCGC 927


RESULT 14
AAV21187
ID   AAV21187 standard; DNA; 53789 BP.
XX
AC   AAV21187;
XX
DT   24-JUL-1998  (first entry)
XX
DE   Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
XX
KW   Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
KW   polyketide synthase; actinomycete; ansamycin; ds.
XX
OS   Amycolatopsis mediterranei.
XX
FH   Key             Location/Qualifiers
FT   CDS             1825..15543
FT                   /*tag=  a
FT                   /label= ORF_A
FT                   /product= "polyketide synthase"
FT   CDS             15550..30759
FT                   /*tag=  b
FT                   /label= ORF_B
FT                   /product= "polyketide synthase"
FT   CDS             30895..36060
FT                   /*tag=  c
FT                   /label= ORF_C
FT                   /product= "polyketide synthase"
FT   CDS             36259..41325
FT                   /*tag=  d
FT                   /label= ORF_D
FT                   /product= "polyketide synthase"
FT   CDS             41373..51614
FT                   /*tag=  e
FT                   /label= ORF_E
FT                   /product= "polyketide synthase"
FT   CDS             51713..5293
FT                   /*tag=  f
FT                   /label= ORF_F
FT                   /product= "polyketide synthase"
XX
PN   WO9807868-A1.
XX
PD   26-FEB-1998.
XX
PF   18-AUG-1997;   97WO-EP04495.
XX
PR   20-AUG-1996;   96EP-0810551.
XX
PA   (NOVS ) NOVARTIS AG.
XX
PI   Engel N,  Schupp T,  Toupet C;
XX
DR   WPI; 1998-169172/15.
DR   P-PSDB; AAW52845-W52850.
XX
PT   Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT   to produce rifamycin and rifamycin analogues
XX
PS   Claim 4; Page 53-102; 205pp; English.
XX
CC   The present sequence represents a Amycolatopsis mediterranei rifamycin
CC   synthesis gene cluster DNA fragment from the present invention. The
CC   DNA fragment comprises a DNA region involved directly or indirectly
CC   in the gene cluster responsible for rifamycin synthesis, including
CC   the adjacent DNA regions to the right and left which, by reason of
CC   their function in connection with rifamycin biosynthesis, qualify
CC   as constituents of this rifamycin gene cluster, and functional
CC   fragments, derivatives or constituents of these. The Amycolatopsis
CC   mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC   for producing rifamycin, rifamycin analogues or precursors. It can also
CC   be used for inactivating or modifying genes involved in ansamycin or
CC   rifamycin biosynthesis. The DNA can be used for constructing mutant
CC   actinomycetes strains from which the natural rifamycin or ansamycin
CC   biosynthesis gene cluster has been partly or completely deleted. The
CC   DNA fragment can be used for assembling a library of polyketide
CC   synthases, which can be used for assembling a library of polyketides.
CC   A hybridisation probe of the invention can be used for identifying DNA
CC   fragments involved in the biosynthesis of ansamycins.
XX
SQ   Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;

  Query Match           38.1%;  Score 351;  DB 19;  Length 53789;
  Best Local Similarity   64.8%;  Pred. No. 1e-50;
  Matches  589;  Conservative    0;  Mismatches  305;  Indels   15;  Gaps    4;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              |||| ||||| ||||| |||||   |||||||||||| |||||      |||||  | |
Db      17257 GTCTGGGTGTTCCCCGGCCAGGGCACGCAGTGGGCCGGCATGGGCCGCGAGCTGCTGGAC 17316

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
              |||||  |||||||||| ||||||   | |||||||   || || | || ||||||
Db      17317 TCCTCGGAGGTGTTCGCCGCGCGGATCGCCGAGTGCGAGACCGCGCTCGGGCGGTGGGTC 17376

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              |||||||  ||   ||| | |||||| | ||      ||  |   ||  ||| |||| |
Db      17377 GACTGGTCGCTGACCGACGTGCTGCGCGGCGA------GGCCGACCTGCTGGACCGGGTC 17430

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              || ||||| || || | |   |||  ||| |||||     | || ||  | |||   |||
Db      17431 GACGTGGTGCAACCGGCGAGCTTCGCCGTGATGGTCGGGCTTGCCGCCGTCTGGGCCTCC 17490

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              |||  |||| ||||| || ||  | ||||| || ||||||||||||||||||||
Db      17491 CTCGGCGTCGAGCCCGAGGCCGTGGTGGGCCACTCGCAGGGCGAGATCGCGGCCGCATGC 17550

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||||  ||||| |||  |||| ||||||||||||||    || ||| |||||||||| ||||
Db      17551 GTGTCCGGGGCACTGTCCCTGGAGGACGCGGCGAAGGTGGTGGCGTTGCGCAGCCAGGCG 17610

Qy        361 CTGGCCGCTGTGCGG---GGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAG 417
              | |||||   || |   |||||||||||||||| || ||  || ||   |  | ||
Db      17611 ATCGCCGCCTCGCTGGCCGGCCGGGGCGGCATGGCGTCGGTCGCGTTGAGCGAAGAAGAC 17670

Qy        418 GTGGAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGC 477
              | |     ||   |   |||| ||||||||| ||  ||  ||| || || ||||||||
Db      17671 GCGACCGCGCGGCT---CGAGCCGTGGGCGGGCCGCGTGGAGGTCGCCGCCGTCAACGGG 17727

Qy        478 CCCCGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGT 537
              ||   |||   |   || | ||||| |||||||| |||||||  |  || |
Db      17728 CCGACGTCCGTGGTGATCGCCGGGGACGCCGAGGCGCTGGACGAAGCCCTCGACGCGCTC 17787

Qy        538 GCCGGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCAT 597
              | || |   || || |||  ||||||| |  |||| ||||| ||||| ||| ||  ||
Db      17788 GACGACCAAGGCGTCCGGATCCGGCGGGTGGCGGTGGACTACGCCTCCCACACCCGGCAC 17847

Qy        598 GTGCAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCC 657
              ||  |  ||  ||| || |   ||   ||| ||||||| |  ||| || ||||| || |
Db      17848 GTCGAAGCCGCGCGCGACGCACTGGCCGAGATGCTGGGCGGGATCCGCGCGCAGGCGCCG 17907

Qy        658 GGCGTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCT---GGACACCACAACCCTGGAC 714
              |  |||||||||| ||| || |||   |||  |||| |     ||| ||       || |||
Db      17908 GAAGTGCCGTTCTACTCGACCGTGACCGGCGGCTGGGTCGAAGACGCCGGCGTGCTCGAC 17967

Qy        715 GCCGCCTACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCC 774
              | || |||||||||||| ||||| | ||  | ||| || ||| ||  || ||     |
Db      17968 GGCGGCTACTGGTACCGGAACCTCCGCCGTCAGGTGCGGTTCGGCCCGGCGGTGGCCGAG 18027

Qy        775 CTGGCGGATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCC 834
              |||   ||  | || ||||| || |||||||| |||||| | || |||   || || |
Db      18028 CTGATCGAGCAGGGCCACCGGGTGTTCGTCGAGGTCAGCGCGCATCCCGTGCTGGTTCAG 18087

Qy        835 GCCATCGAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGC 894
```

```
Db      18088 CCGATCAACGAACTCGTCGACGACACCGAAGCCGTGGTCACCGGGACGCTGCGGCGCGAG 18147

Qy        895 CGCGGCGAC 903
              ||||| |
Db      18148 GACGGCGGC 18156
```

RESULT 15
AAF88335
ID AAF88335 standard; DNA; 7788 BP.
XX
AC AAF88335;
XX
DT 28-AUG-2001 (first entry)
XX
DE S. spinosa DNA fragment encoding ORF18, SEQ ID 41.
XX
KW Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; polyketide synthase; ds.
XX
OS Saccharopolyspora spinosa.
XX
PN DE19957268-A1.
XX
PD 08-MAR-2001.
XX
PF 29-NOV-1999; 99DE-1057268.
XX
PR 27-AUG-1999; 99DE-1040596.
XX
PA (FARB) BAYER AG.
XX
PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
XX
DR WPI; 2001-267102/28.
DR P-PSDB; AAB70965.
XX
PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives
PT -
XX
PS Claim 7a; Page 165-178; 354pp; German.
XX
CC This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (II); (ii) to generate a library of polyketide synthases; (iii) for
CC adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC aglycone; and (iv) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC macrolides with insecticidal, but not antibacterial, activity, and can
CC also be used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (II) at significantly increased levels or produce new derivatives of
CC (II). This sequence, ORF 18, encodes an S. spinosa polyketide synthase.
XX
SQ Sequence 7788 BP; 1127 A; 2240 C; 2936 G; 1485 T; 0 other;

```
  Query Match           38.0%;  Score 349.6;  DB 22;  Length 7788;
  Best Local Similarity  62.9%;  Pred. No. 1.9e-50;
  Matches  576;  Conservative    0;  Mismatches  334;  Indels    6;  Gaps    2;

Qy          1 GTCTTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCC 60
              |||||||| ||||||||||||||  | ||||||  ||||||||||  ||  |||||||
Db       1579 GTCTTCGTCTTTCCCGGACAGGGCAGCCAGTGGATCGGGATGGGTAGGCAGCTGCTCGAA 1638

Qy         61 TCCTCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTG 120
```

```
Db       1639 ACCTCCGAGGTTTTCGCGGCGTCGATGTCGGACTGCGCCGACGCATTGGCGCCGCACCTG 1698

Qy        121 GACTGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCC 180
              || |||||  || ||||  | ||||||    |  || || | ||||       || |  | |
Db       1699 GATTGGTCCCTGCTGGATGTGCTGCGCAACGCGGCCGGCGCTGCGCACCTTGACCACGAC 1758

Qy        181 GATGTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCC 240
              ||||| |||||||| | |||||||  | |||||||  ||   | ||  |  || ||||||||
Db       1759 GATGTCGTCCAGCCCGCGCTGTTCGCCATCATGGTCTCGCTCGCGGAGCTCTGGCGTTCG 1818

Qy        241 TACGGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCAT 300
              |  ||  |   || |  |||||| | || || || ||||| ||||||||||| ||
Db       1819 TGGGGCGTGCGTCCGGTGGCGGTCGTCGGGCACTCGCAGGGGGAGATCGCGGCGGCCTGC 1878

Qy        301 GTGTGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCG 360
              ||    ||||| |||  | |    || || || | |   || ||| |||||||| ||
Db       1879 GTCGCCGGGGCCCTGTCCGTCCGCGATGCCGCCAGGGTGGTGGCGGTGCGCAGCAGGCTT 1938

Qy        361 CTGGCCGCTGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 420
              ||| | ||  ||  ||| | |||| |||||||| ||| ||  || ||| | ||||||
Db       1939 CTGACGGCGCTGGCCGGCAGTGGCGCGATGGCCTCGTTGCAGCATCCCGCCGAAGAGGTG 1998

Qy        421 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 480
              |||  || |   ||  |||  |  ||| |  | ||||||| ||| ||||| ||
Db       1999 CGGCAAATCCT---GTTGCCCTGGCGCGATCGGATCGGCGTGGCGGGGGTGAACGGACCG 2055

Qy        481 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 540
              || |||   || || |||||   |||||| |||  ||  |||||||  | || ||
Db       2056 TCGTCGACCCTGGTGTCAGGGGACCGGGAGGCGATGGCGGAACTGCTGGCCGAGTGCGCA 2115

Qy        541 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 600
              | |   | | | |||   || ||||| || || || || ||||| || |  || ||  |
Db       2116 GACCGAGAGCTCCGGATGCGCCGGATTCCCGTTGAATACGCCTCCCATTCGCCTCACATC 2175

Qy        601 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 660
              ||   | ||||| ||| |||||| |||| ||| |    ||   ||   | ||    ||
Db       2176 GAGGTTGTCCGGGATGAGCTGCTGGGGCTGTTGGCGCCGGTCGAACCCAGGACGGGAAGC 2235

Qy        661 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 720
              | ||| |||  || ||| |   ||   | |||||||    | |  |||||||| | |
Db       2236 ATCCCGATCTATTCGACGACGACCGGGGACCTGCTGGACCGGCCGA---TGGACGCCGAC 2292

Qy        721 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 780
              ||||||||||||||||  || ||||| |  |||   | |||||| |||||||  |
Db       2293 TACTGGTACCGCAACCTTCGTCAACCGGTGCTGTTCGAAGCGGCCGTCGAGGCCCTGTTG 2352

Qy        781 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 840
              |   ||  ||  ||  ||| ||||  ||||||| |||||    ||     |   ||||
Db       2353 AAGCGGGGGTACGACGCATTCATCGAGATCAGCCCACACCCGGTGCTGACTGCGAACATC 2412

Qy        841 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 900
              | || ||| |||       | |     || ||    ||| |||| |  ||||||||||
Db       2413 CAGGAAACCGCCGTGCGAGCAGGGCGGGAGGTAGTGGCGCTCGGGACACTCCGCCGCGGC 2472

Qy        901 GACAACGACACCCGCC 916
              ||    | ||  || |
Db       2473 GAAGGTGGCATGCGGC 2488

Search completed: June 17, 2003, 13:48:25
Job time : 194.303 secs
```

THIS PAGE BLANK (USPTO)

```
Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCG 1299
              ||| | ||||| || |||||  ||||  | ||  |  | |||
Db       6586 ATCAGTGGCACCAACGCGCACCTGATCCTTGAACAACCTCCG 6627
```

Search completed: June 17, 2003, 13:48:14
Job time : 269.094 secs

RESULT 15
AAF88338
ID AAF88338 standard; DNA; 14775 BP.
XX
AC AAF88338;
XX
DT 28-AUG-2001 (first entry)
XX
DE S. spinosa DNA fragment encoding ORF21, SEQ ID 47.
XX
KW Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; polyketide synthase; ds.
XX
OS Saccharopolyspora spinosa.
XX
PN DE19957268-A1.
XX
PD 08-MAR-2001.
XX
PF 29-NOV-1999; 99DE-1057268.
XX
PR 27-AUG-1999; 99DE-1040596.
XX
PA (FARB) BAYER AG.
XX
PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
XX
DR WPI; 2001-267102/28.
DR P-PSDB; AAB70968.
XX
PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives
PT -
XX
PS Claim 7a; Page 239-264; 354pp; German.
XX
CC This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (II); (ii) to generate a library of polyketide synthases; (iii) for
CC adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC aglycone; and (iv) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC macrolides with insecticidal, but not antibacterial, activity, and can
CC also be used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (II) at significantly increased levels or produce new derivatives of
CC (II). This sequence, ORF 21, encodes an S. spinosa polyketide synthase.
XX
SQ Sequence 14775 BP; 1969 A; 4181 C; 5731 G; 2894 T; 0 other;

Query Match 50.0%; Score 650.4; DB 22; Length 14775;
Best Local Similarity 70.5%; Pred. No. 7.1e-103;
Matches 918; Conservative 0; Mismatches 366; Indels 18; Gaps 3;

```
FT                   /product= "polyketide synthase"
FT   CDS             36259..41325
FT                   /*tag=  d
FT                   /label= ORF_D
FT                   /product= "polyketide synthase"
FT   CDS             41373..51614
FT                   /*tag=  e
FT                   /label= ORF_E
FT                   /product= "polyketide synthase"
FT   CDS             51713..5293
FT                   /*tag=  f
FT                   /label= ORF_F
FT                   /product= "polyketide synthase"
XX
PN   WO9807868-A1.
XX
PD   26-FEB-1998.
XX
PF   18-AUG-1997;   97WO-EP04495.
XX
PR   20-AUG-1996;   96EP-0810551.
XX
PA   (NOVS ) NOVARTIS AG.
XX
PI   Engel N,  Schupp T,  Toupet C;
XX
DR   WPI; 1998-169172/15.
DR   P-PSDB; AAW52845-W52850.
XX
PT   Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT   to produce rifamycin and rifamycin analogues
XX
PS   Claim 4; Page 53-102; 205pp; English.
XX
CC   The present sequence represents a Amycolatopsis mediterranei rifamycin
CC   synthesis gene cluster DNA fragment from the present invention. The
CC   DNA fragment comprises a DNA region involved directly or indirectly
CC   in the gene cluster responsible for rifamycin synthesis, including
CC   the adjacent DNA regions to the right and left which, by reason of
CC   their function in connection with rifamycin biosynthesis, qualify
CC   as constituents of this rifamycin gene cluster, and functional
CC   fragments, derivatives or constituents of these. The Amycolatopsis
CC   mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC   for producing rifamycin, rifamycin analogues or precursors. It can also
CC   be used for inactivating or modifying genes involved in ansamycin or
CC   rifamycin biosynthesis. The DNA can be used for constructing mutant
CC   actinomycetes strains from which the natural rifamycin or ansamycin
CC   biosynthesis gene cluster has been partly or completely deleted. The
CC   DNA fragment can be used for assembling a library of polyketide
CC   synthases, which can be used for assembling a library of polyketides.
CC   A hybridisation probe of the invention can be used for identifying DNA
CC   fragments involved in the biosynthesis of ansamycins.
XX
SQ   Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;

  Query Match           50.3%;  Score 655;  DB 19;  Length 53789;
  Best Local Similarity  70.9%;  Pred. No. 1.1e-103;
  Matches  922;  Conservative    0;  Mismatches  355;  Indels   24;  Gaps    3;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || |||||||| ||| |||  ||||| ||||| ||||| || || || |    |||  ||
Db       7918 GAGCCGATCGCGATCGTCGCGATGGCGTGCCGCTTCCCCGGTGGCGTGCACAGCCCCGAG 7977

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              |||||||||  ||||| ||||   ||||  |||||| ||     ||||||| ||||||||
Db       7978 GACCTGTGGCGGCTGGTCGCCGACGGCGCCGACGCCGTCACCGAGTTCCCCGCCGACCGC 8037

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||  ||| || | ||   ||  | | ||||||   ||    ||||||   ||
Db       8038 GGCTGGGACACCGACCGGCTCTACCACGAAGACCCCGACCACGAAGGCACGACGTACGTC 8097

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              | |   ||||  |||||  ||||||| ||    |||||||||| ||||||||| |||

Db       8098 CGGCACGGCGCCTTCCTCGACGACGCCGCCGGGTTCGACGCCGCCTTCTTCGGCATCTCG 8157

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              ||    ||||| || || ||||| ||||||||||||||||||||||||||||||||||||
Db       8158 CCGAACGAGGCGCTCGCCATGGACCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 8217

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||||||||||||| | ||||| |||    || |  |||       | |||||||||
Db       8218 CTGTTCGAGCGGGCCGCGATCGACCCGACCACGCTGGCCGGCCAGGACATCGGCGTCTTC 8277

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              || |||| |  |  | ||||||| ||  ||| ||| ||                 | |||
Db       8278 GCCGGCGTCAACAGCCACGACTACAGCATGCGGATGCAC--------------CGGGCC 8322

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              || |  || ||||||  |   ||||||||     ||| |||||| || || |||||  ||
Db       8323 GCCGGTGTCGAGGGCTTCCGGCTCACCGGCGGTTCGGCCAGCGTGCTCTCCGGCCGCGTC 8382

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              ||||||  ||||||  | |||||  |||||  ||||  || |||||||  |||||| | |||
Db       8383 GCCTACCACTTCGGCGTCGAAGGCCCGGCCGTCACGGTCGACACGGCCTGCTCGTCTTCG 8442

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              || ||  |||||||  |||||   ||| | |||| |   || |||||| |  | |||||
Db       8443 CTGGTCGCGCTGCACATGGCGGTGCAGGCCCTGCAGCGCGGCGAGTGCTCCATGGCGCTC 8502

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              || ||||||||   ||| |||  |||  ||| |||| |||||||||| |||||||||
Db       8503 GCGGGCGGCGTGATGGTGATGGGCACGGTCGAGACGTTCGTCGAGTTCTCGCGGCAGCGC 8562

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              ||||||  |  |||||| | |||||||||| | | |  | ||  ||||||||||||||||
Db       8563 GGGCTGGCCCCCGACGGCCGCTGCAAGGCGTTCGCCGACGGCGCGGACGGCACCGGCTGG 8622

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              |||||||||||||| ||||  |||||||||| |||| || ||   ||| |  |||||
Db       8623 TCCGAGGGCGTCGGGCTGCTCCTGGTGGAGCGGCTGTCCGAGGCTCAGCGTCGCGGGCAC 8682

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              | ||| || || ||||| |||||    |||||||||   |||||  |||||||| ||  ||
Db       8683 CAGGTCCTCGCCGTGGTCCGCGGGTCGGCGGTCAACTCCGACGGCGCGTCGAACGGCTTG 8742

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||| ||||||||  ||| | ||| |||| ||||||||  ||||  ||||    ||| ||
Db       8743 ACGGCCCCGAACGGCCCGTCCCAGCAGCGCGTGATCCGCAAGGCACTGGCCGCCGCCGGA 8802

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              |||||    |||| || || | |||||||| |||||| || || ||||| ||||| ||
Db       8803 CTGTCCACATCGGACGTCGACGCGGTGGAGGCGCACGGCACCGGGACGACCCTGGGCGAC 8862

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              |||||||||||  ||||| |||| ||||| ||||| ||   |  | | || | |||||||
Db       8863 CCGATCGAGGCCGAGGCGCTGCTGGCCACCTACGGCCA---GAACCGGGAAACGCCGCTG 8919

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              ||||| |||||  ||||||| ||| ||||||||||   ||| || ||||||| |  ||
Db       8920 TGGCTCGGGTCGGTGAAGTCGAACCTCGGGCACACGCAGGCGGCTGCGGGTGTCGCAGGC 8979

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              || |||||||||| ||||  ||||  | || ||  |||| |||||| |||| || ||
Db       8980 GTGATCAAGATGGTCATGGCCATGCGCCACGGCGTCCTGCCCCGGACGCTGCACGTCGAC 9039

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              |||||||  |  | ||||||||||  || ||  |||||   ||||||||  |||||  |
Db       9040 CGGCCGTCGTCCTATGTGGACTGGTCGGCCGGTGCGGTCGAGCTGCTGACCGAGGCACGG 9099

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              |||  | |  ||      || |   ||| || |||||  ||||||||||||||  |||
Db       9100 GACTGGGTGAGCAAC------GGCCACCCGCGCCGCGCGGGCGTGTCGTCGTTCGGCATC 9153

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGC 1301
              |||||||| || ||||| ||   |  | || ||||   | ||
Db       9154 GGCGGCACCAACGCGCACGTCGTCCTCGAAGAGGTTGCCGC 9194
```

```
XX
PA   (STAM ) DSM NV.
XX
PI   Martin JF,  Aparicio JF,  Colina AJ;
XX
DR   WPI; 2001-080693/09.
DR   P-PSDB; AAB31557.
XX
PT   New polynucleotides encoding enzymes involved in the biosynthesis of
PT   pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT   the synthesis of new compounds  -
XX
PS   Disclosure; Page 39-46; 116pp; English.
XX
CC   The present sequence encodes a polyketide synthase which is associated
CC   with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC   is useful for the oxidative modification of a methyl group of a suitable
CC   compound, e.g. a bioactive compound including a secondary metabolite,
CC   antibiotics and anticancer agents. Recombinant cells comprising the
CC   gene are useful for the production of pimaricin. The polyketide synthase
CC   polynucleotide may be over expressed in Streptomyces, leading to an
CC   increase in the biosynthesis of pimaricin, as a source of primers for
CC   amplification reaction and as probes.
XX
SQ   Sequence 5544 BP; 699 A; 2052 C; 2023 G; 770 T; 0 other;

  Query Match             50.9%;  Score 662.6;  DB 22;  Length 5544;
  Best Local Similarity   71.1%;  Pred. No. 6.2e-105;
  Matches  926;  Conservative    0;  Mismatches  359;  Indels   18;  Gaps    3;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              |||||| | || ||| | |||||||| |||||  | ||||||||| |    |||||  |
Db       1981 GACCCGGTGGCGATCGTGGGCATGGCCTGCCGGCTGCCGGGCGGAGCCGACTCCCCCGAC 2040

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              | |||||||||||||  ||||   ||    |||||  ||    || |||||||| ||||||
Db       2041 GCCCTGTGGGAGCTGCTCGCCGACGGGACCGACGCGATGTCCCCCTTCCCCACGGACCGC 2100

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||  |  ||| | ||   || | ||||  ||   |  || || || |  ||  |
Db       2101 GGCTGGGACCTGGACCGGCTGTTCGACGAGGATGCCGACCGCCCGGGTACCTCGTACGCC 2160

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              |  | |||||| ||||| |||||||||||   |||||||||| ||||||||||  |
Db       2161 CGCGAAGGCGGCTTCCTGCACGACGCGGGCGACTTCGACGCGGGCTTCTTCGGCCTGTCG 2220

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              |  |||||   |||||  |||||||||||||||||||| |||||| || ||||||||
Db       2221 GACCAGGAGGCGACGGCGACCGATCCGCAGCAGCGGCTGCTTCTGGAGGCGGCCTGGGAG 2280

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              | |||||||||||||| ||||| |||| |||  |  | ||   |||||||||||| |||
Db       2281 ACCTTCGAGCGGGCGGGCATCGACCCGCAGTCCCTGAGGGGAAGCCGTACGGGCGTGTTC 2340

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              |||||||  |   |  || ||| ||  |  |   |  | |||||  |    |   |
Db       2341 ACGGGCGCGATGGACCGCGGCTATGGAACCAGCGCGTCCGCCGCGCCCAGCGCATG---- 2396

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
                 |||| ||    || ||||||| ||| | |||||||   | ||||| || |||
Db       2397 --------GGAGAGCATGCTCATCACCGGGACCGCCGGCAGCGCGGTCTCGGGGCGCATC 2448

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              ||||||| || |||||| ||||| || ||| | || || |||||  |  | |||| | ||
Db       2449 GCCTACACCTACGGGCTCGAAGGCCCCGCGCTGACGGTCGACACCGCCTCCTCGTCCTCC 2508

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              |||||  | |||||||||| |||| |||||||| ||||| |||  |   || ||||||
Db       2509 CTCGTCGCCCTGCATCTGGCCTGCCGGTCGCTGCGCTCGGGCGAGACCGACCTGGCGCTG 2568

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              |||||||||||| | |||||| | |||| |     |||    | |||||||||| ||||
Db       2569 GCCGGCGGCGTCACCGTCATGGCGACCCCGGCGCCCTTCGCACACTTCTCCCGGCTGCGC 2628

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              | ||||||    |||  | | | || ||||||| |||  || ||  ||||| | |  |||
Db       2629 GCGCTGTCCCCCGACTCCCGCTCCATGGCGTACGCGGACGCCGCGAACGGCTCGGCGTGG 2688

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
                 |||||||  ||| ||||  || ||||||| ||    || ||   |||    || ||
Db       2689 TCGGAGGGCGCGGGGCTGCTTCTGCTGGAGCGGCTGAGCGACGCCCGGCGCAACGGACAC 2748

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              || || |||||| | ||||| |||    || || || |||||||| || || || |||||
Db       2749 CGTGTCCTGGCGCTCGTACGGGGCTCCGCCGTGAATCAGGACGGCGCCTCCAACGGGCTC 2808

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              || ||||||| ||| || || ||| |||| || ||||| |||||  ||||  |||| |||
Db       2809 ACCGCGCCGAGCGGACCCGCACAGCAGCGCGTCATCCGCCAGGCCCTGGCCGACGCCGGG 2868

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              || |   |  ||| ||||| |  |||||||||||||| || |||||| |||| || ||
Db       2869 CTGACGCCGCAGGACGTGGACGCCGTGGAGGGGCACGGCACCGGCACGCCGCTCGGCGAC 2928

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGG---GCAGCGGGCCGGTGACAGGCCG 1017
              || |||||||| |||||| |||| ||||||||||    ||||||  | |  ||  |||||
Db       2929 CCCATCGAGGCGCAGGCGCTGCTGGCCACGTACGGCCAACAGCGGCCTGTGGAACGGCCG 2988

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077
              ||||||||||||  |||||| ||| |||||||||    || ||||| || || |
Db       2989 TTGTGGCTGGGGTCGGTGAAGTCGAACTTCGGGCACACACAAGCCGCCGCCGGGGTCGTC 3048

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137
              || |||||||||| |||| | ||| ||||  | || ||  |||||| |||| |||| |||
Db       3049 GGCGTCATCAAGACGGTGCTCGCGCTGCGCCACGGCGTCCTGCCGCAGACGCTGCACGTG 3108

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197
              ||    || ||| |  |||||||||||||||| ||  |||| |||||||||| ||||||
Db       3109 GACGCTCCCTCGGCCAAGGTGGACTGGTCCGCCGGTTCGGTACGGCTGCTGACCGAGGCG 3168

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257
                 ||| |||||  ||||   |   || ||   ||| |||||||| ||||| |||||||||
Db       3169 CGGCCCTGGCCACGGGA---GAGCGGACGTACGCGCCGGGCGGGGGTGTCCTCGTTCGGG 3225

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGG 1300
              ||  |||||| || ||||| |||||  ||||||||||||||
Db       3226 CTCACCGGCACCAACGCGCACGTGATCCTGGAGGAGGCGCCGG 3268


RESULT 14
AAV21187
ID   AAV21187 standard; DNA; 53789 BP.
XX
AC   AAV21187;
XX
DT   24-JUL-1998  (first entry)
XX
DE   Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
XX
KW   Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
KW   polyketide synthase; actinomycete; ansamycin; ds.
XX
OS   Amycolatopsis mediterranei.
XX
FH   Key             Location/Qualifiers
FT   CDS             1825..15543
FT                   /*tag=  a
FT                   /label= ORF_A
FT                   /product= "polyketide synthase"
FT   CDS             15550..30759
FT                   /*tag=  b
FT                   /label= ORF_B
FT                   /product= "polyketide synthase"
FT   CDS             30895..36060
FT                   /*tag=  c
FT                   /label= ORF_C
```

```
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22601-W22605.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Claim 2; Pages 8-66; 220pp; English.
XX
CC   This sequence represents the tylactone synthase gene cluster of the
CC   invention. This sequence is also referred to as the tylG gene, and was
CC   isolated from Streptomyces fradiae. This sequence encodes multifunctional
CC   proteins which direct the synthesis of the polyketide tylactone, isolated
CC   from Streptomyces fradiae. Tylactone is the basic building block of the
CC   antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC   type of carboxylic acids incorporated, the number of carboxylic acids
CC   incorporated and/or the post-condensation reactions performed, thereby
CC   resulting in novel tylosin-related polyketides.
XX
SQ   Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;

  Query Match             52.0%;  Score 676.4;  DB 18;  Length 43280;
  Best Local Similarity   72.0%;  Pred. No. 2.4e-107;
  Matches  932;  Conservative    0;  Mismatches  341;  Indels   21;  Gaps    3;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || |||||||||| || |||||||||| ||||| | ||| || || |       ||||  |
Db      31337 GAGCCGATCGCCGTCGTCGGCATGGCCTGCCGCTACCCCGGAGGCGCGAACACCCCCGAA 31396

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
               |  | ||||| |||  || | | |||   ||||||  || |||| | ||| |||||||
Db      31397 CAGTTCTGGGAACTGCTCGACACCGGCACCGACGCCGCCGCGCCGATGCCCTCCGACCGG 31456

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              ||||||  |||  |   ||   ||| ||  |||||  ||    | ||      |||
Db      31457 GGATGGGACACCCACGGGCTGTACGACCCCGACCCGGCGGCAGCGGGGCGCACCTACTGC 31516

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              | ||  ||||| |||||  ||||||||||   |||||||||| | |||||||||  ||
Db      31517 CGGGAGGGCGGCTTCCTCCACGACGCGGGCGACTTCGACGCGGACTTCTTCGGGATTTCG 31576

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              || ||||||||  |||||||||||||||||||||||||| |||||||||| |||||||||
Db      31577 CCGCGTGAGGCGGTGGCGATGGATCCGCAGCAGCGGCTGTTGCTGGAGACCTCCTGGGAG 31636

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||| ||||   ||| || ||||| ||||      | |||||   ||| || || || | |
Db      31637 GCGATCGAAGCGGCCGGTATCGACCCGCGAGGACTCCGCGGCAGCCGCACCGGGGTGTAC 31696

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              |||||||      | || ||||  |   |     | || |  | |||
Db      31697 GTGGGCGCCTGGGACAGCGGCTACACCGGCCAGGCGCACGCGCCCTCG---------- 31744

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              || ||  |||||| | || | || |||||    |   | ||||  ||   || |||||||||
Db      31745 GCCGAGTTGGAGGCCGACCTGCTGACCGGCGGCGTCGTCAGCTTCACCTCCGGCCGTATC 31804

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              |||||||   | ||||||| || |||||  | ||||||||||| | |||  ||| | |||
Db      31805 GCCTACACGCTGGGGCTGGAGGGTCCGGCCTTGACCGTGGACACCGCGTGTTCGTCGTCG 31864

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              || ||  |.|||||    |||  ||| ||||||||   || || |||   || ||| ||
Db      31865 CTGGTCGCCCTGCACAACGCGGCGCAGGCGCTGCGGCGCGGCGAATGCGACCTGGCGTTG 31924

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              ||||| || ||  |||| ||| | ||||  |   ||||| || ||||| ||||||||||
Db      31925 GCCGGTGGTGTGACGGTGATGGCGACCCCGGCGGTGTTCGTCCAGTTCGCCCGGCAGCGA 31984

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              |||||| ||  ||||||| | |||||||||| | | |  || |||||||||  ||||
Db      31985 GGGCTGGCGCCGGACGGCCGCTGCAAGGCGTTCGCCGACGCCGCCGACGGCTTCGGCCCC 32044

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              | |||||| || |||||| || ||||||||||||||||||||||   |||| ||||||||
Db      32045 GCCGAGGGTGTGGGGATGGTGCTGGTGGAGCGGTTGTCGGATGCCCGGCGGTTGGGGCAT 32104

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              | |||| ||||||||||  | ||     ||||| || ||||||||||||||||||  ||
Db      32105 CCGGTGTTGGCGGTGGTGTGTGGGTCGGCGGTGAATCAGGACGGTGCGTCGAATGGTTTG 32164

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||||||||  |||||| | ||||||||||||||  || ||||||||||  ||| ||| ||
Db      32165 ACGGCGCCGAGTGGTCCGTCGCAGGAGCGGGTGATTCGTCAGGCGTTGGGGAATGCGCGG 32224

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||| | |||||||||||||||||||||||||| ||| ||||||||  |||  ||||||||||
Db      32225 TTGACGGTGGCGGATGTGGATGTGGTGGAGGCGCATGGGACGGGGACGCGGCTGGGTGAT 32284

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGC---AGCGGGCCGGTGACAGGCCG 1017
              |||||||||||  |||||||||||  | |||||  ||||    | ||  |  ||||   | |||
Db      32285 CCGATCGAGGCGCAGGCGTTGCTGGGGACGTATGGGCGGGATCGTGATGGTGGGCGTCCG 32344

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077
              |||||  |||||||  |||||||  || || ||  ||  |   |||  ||  |||||  |||| |
Db      32345 GTGTGGTTGGGGTCGTTGAAGTCGAATATTGGTCATGCTCAGGCGGCTGCGGGGGTGGCT 32404

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137
              || || |||||||||||| |||||  |||||  | |||  |||||||||||||||||||||
Db      32405 GGTGTGATCAAGATGGTGTTGGCGATGCGGTATGGGTGGTTGCCGCGGACGTTGCATGTG 32464

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197
              ||| |||||   | ||| |||||||||||| || || |  ||||||  |||||||  |||||
Db      32465 GATGAGCCGAGCCGGCATGTGGACTGGTCGGCTGGTGGTGTGCGGTTGCTGACCGAGGCG 32524

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257
                 |  |||||||||       ||  |  |||  ||| |||||||  || ||  | || ||
Db      32525 CGGGAGTGGCCGGG------GGTGGACCGGCCGCGTCGGGCGGCGGTCTCCGCCTTTGGT 32578

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGG 1291
              || | ||  ||  ||  ||  ||| ||||   | || |
Db      32579 GTCAGTGGTACCAACGCCCATCTGATCCTCGAAG 32612


RESULT 13
AAF24891
ID   AAF24891 standard; DNA; 5544 BP.
XX
AC   AAF24891;
XX
DT   20-APR-2001  (first entry)
XX
DE   Pimaricin biosynthesis associated polyketide synthase gene.
XX
KW   Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW   anticancer; pimaricin; ss.
XX
OS   Streptomyces natalensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..5544
FT                   /*tag= a
FT                   /product= "polyketide synthase"
XX
PN   WO200077222-A1.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-EP06227.
XX
PR   14-JUN-1999;   99EP-0201893.
```

```
CC   particular, the inactivation is mediated by a change in the KS domain
CC   that renders it incapable of binding substrate (the KS1-o mutation),
CC   rendered by mutation in the codon for the active site cysteine. The
CC   oleandolide PKS is useful for synthesizing polyketides, which are useful
CC   as antibiotics and motilides. Heterologous expression of oleandolide PKS
CC   in host cells such as Streptomyces coelicolor and S. lividans is also
CC   made possible. Unmodified oleandolide compounds can be provided to
CC   cultures of Saccharopolyspora erythraea and converted to the
CC   corresponding derivatives of erythromycins A-D.
XX
SQ   Sequence 50937 BP; 6672 A; 16253 C; 19272 G; 8740 T; 0 other;

  Query Match             52.9%;  Score 689;  DB 21;  Length 50937;
  Best Local Similarity   72.3%;  Pred. No. 1.7e-109;
  Matches  944;  Conservative    0;  Mismatches  340;  Indels   21;  Gaps    3;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || |||||||| ||| | || |||||||||| |||||||| ||  ||   |||||| ||
Db      13407 GAGCCGATCGCGATCGTGGGAATGGCATGCCGCTTCCCGGGAGGCATCGCTTCCCCGGAG 13466

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              || ||||||||  ||  | || | || |  ||||   ||       | ||      |||||
Db      13467 GATCTGTGGGACGTGCTCACCGCTGGTGAGGACGTTGTCTCCGGACTGCCGCAGAACCGC 13526

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||  |  ||  |  |   |||  |  ||  ||    | | ||||  |||  ||   |
Db      13527 GGGTGGGACTTGGGGCGCCTGTACGATCCCGATCCGGACCGGGCCGGTACGTCATACATG 13586

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              |  |  || |  ||||| |||||.|||| | | |||||||| | |||||||| |||
Db      13587 CGTGAGGGTGCTTTCCTGCACGAGGCGGGGGAGTTCGACGCGGCCTTCTTCGGTATCTCG 13646

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              ||.||||||||  |||||||||| |||||||||||| ||||||||||||||||||||||
Db      13647 CCGCGTGAGGCGTTGGCGATGGACCCGCAGCAGCGGTTGCTGCTGGAGACGTCCTGGGAG 13706

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||  |||| ||||| || |||  |||       | |     ||     |  || || || |||
Db      13707 GCCCTCGAACGGGCCGGCATCACTCCTTCCAAGCTGGCGGGCAGTCCGACCGGTGTGTTC 13766

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
                 |||         |   ||||||| | | |     || |  || | |||| |||
Db      13767 TTCGGCATGTCGAACCAGGACTACGCCGCCCAGGCGGGCGACGTGCCGTCCGAG------ 13820

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
                 |||||||| || | ||||||||  |||    |||||||   ||||| ||| | |
Db      13821 ------CTGGAGGGCTACCTGCTCACCGGCTCCATCTCCAGCGTCGCTTCGGGGCGTGTT 13874

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              || ||||  ||||| || || ||||| ||| | || ||||| |||| ||| ||| | |||
Db      13875 GCTTACACGTTCGGTCTTGAGGGGCCTGCGGTGACGGTGGATACGGCGTGTTCGTCGTCG 13934

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              | ||| || ||||||||||    |||  | |||||  |||||||||  |||| ||| ||
Db      13935 TTGGTGGCGTTGCATCTGGCGGTGCAGGGGTTGCGGCGGGGTGAGTGTTCGCTTGCGTTG 13994

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              |  || || ||  |||| |||||  | |  |  | ||| |    ||||   ||||||||
Db      13995 GTGGGTGGGGTGACGGTGATGTCGTCGCCGGTGACGTTGACGACGTTCAGTCGGCAGCGG 14054

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              ||  |||||||||| ||  |||||||||||| | ||||| | || || ||    ||
Db      14055 GGTTTGTCGGTGGATGGGCGGTGCAAGGCGTTCGCGGCTTCGGCGGATGGTTTTGGTGCT 14114

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              | |||||| || ||  || |||||||||||||||||||||||||||  |||| |||| |||
Db      14115 GCCGAGGGTGTGGGTGTGTTGTTGGTGGAGCGGTTGTCGGATGCGCGGCGGTTGGGTCAT 14174

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              |||||| |||||||||| || || |||||||||| ||||| |||||||| ||||| |||
Db      14175 CGGGTGTTGGCGGTGGTGCGGGGGAGTGCGGTCAATCAGGATGGTGCGTCCAATGGTCTG 14234

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
                  |||||||||| |||||| | ||| ||||||||||||   |||||||| |||||||
Db      14235 GCGGCGCCGAATGGTCCGTCGCAGCAGCGGGTGATCCGTGCGGCGTTGGCTGACGCGGGT 14294

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              || |    || |||||||||||||||||| ||| || ||||| |||  | ||||||||
Db      14295 CTGGCTCCTGCCGATGTGGATGTGGTGGAGGCGCATGGCACGGGGACGCGGTTGGGTGAT 14354

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAG---CGGGCCGGTGACAGGCCG 1017
              |||||||||| |||||||||| || ||||| ||||||   ||  || |||   | |||
Db      14355 CCGATCGAGGCTCAGGCGTTGCTGGCGACGTATGGGCAGGGTCGTACCAGTGGGCGTCCG 14414

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077
              |||||||||||| |||||| |||||||||| ||   ||| || || |||||||  |
Db      14415 GTGTGGCTGGGGTCGGTGAAGTCGAACATCGGGCATACGCAGGCGGCGGCCGGTGTGGCT 14474

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137
              || || || |||||||| |||||||| |   ||| ||| |||||  ||||||||||||
Db      14475 GGTGTGATGAAGATGGTGCTGGCGTTGGGTCGGGGTGTGGTGCCGAAGACGTTGCATGTG 14534

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197
              ||| ||||||| ||||| |||||||||| || || | ||||  | ||  |  ||||||
Db      14535 GATGAGCCGTCACCGCATGTGGACTGGTCGGCTGGTGAGGTGGAGTTGGCGGTTGAGGCG 14594

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257
              ||||||||| || |      ||  |||||| |||||||||| || ||||||||||||||
Db      14595 GTGCCGTGGTCGCG------GGGTGGGCGGGTGCGGCGGGCTGGTGTGTCGTCGTTCGGG 14648

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              ||| |||||||||||||||||||||  |||||||||||| |||
Db      14649 ATCAGCGGCACGAATGCGCATGTGATCGTGGAGGAGGCGCCTGCG 14693


RESULT 12
AAT80413
ID   AAT80413 standard; DNA; 43280 BP.
XX
AC   AAT80413;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Tylactone synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             816..14243
FT                   /*tag=  a
FT                   /transl_except= (pos: 816..818, aa: Met)
FT                   /note= "ORF1 encodes protein shown in AAW22601"
FT   CDS             14351..19945
FT                   /*tag=  b
FT                   /transl_except= (pos: 14351..14353, aa: Met)
FT                   /note= "ORF2 encodes protein shown in AAW22602"
FT   CDS             20010..31199
FT                   /*tag=  c
FT                   /transl_except= (pos: 20010..20012, aa: Met)
FT                   /note= "ORF3 encodes protein shown in AAW22603"
FT   CDS             31232..36067
FT                   /*tag=  d
FT                   /note= "ORF4 encodes protein shown in AAW22604"
FT   CDS             36249..41774
FT                   /*tag=  e
FT                   /note= "ORF5 encodes protein shown in AAW22605"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
```

```
Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              |.|||||.||.||..|||||.||||||||||.|.||.||.|||...||....||.|||
Db       8978 GCGGAGGGTGTGGGTGTGCTGCTGGTGGAGCGGCTTTCCGACGCGCGCCGCAACGGTCAT 9037

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              ||||||||||||||||.||.||||||||||||.||||||||||...||||||||
Db       9038 CGGGTGCTGGCGGTGGTGCGGGGCAGTGCGGTCAATCAGGACGGTGCGAGCAATGGGCTG 9097

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||||||||..||||||||.|||.||||||||||||..|||||.||||..|.||||||
Db       9098 ACGGCGCCGAGTGGTCCGGCGCAGCAGCGGGTGATCCGTGAGGCGCTGGCTGATGCGGGG 9157

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
               ||      ||.||.||||||||||||||||.||||||.|||||.|||.||||||||||
Db       9158 CTGGTGCCCGCCGACGTGGATGTGGTGGAGGCGCACGGTACGGGGACGGCGCTGGGTGAT 9217

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              ||||||||||...|||.||||.||||||||||||.|...|...|.|.|...|||.||
Db       9218 CCGATCGAGGCGGGTGCGCTGCTGGCCACGTACGGGCGGGAGCGGGTCGGCGATCCGTTG 9277

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              |||||.|||||..|||||||.||||||||||..|...|||.||.||||||||||||||
Db       9278 TGGCTCGGGTCGTTGAAGTCGAACATCGGGCATGCGCAGGCGGCTGCGGGTGTGGGTGGT 9337

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              ||.||||||.|||||..||.|.|||||.|.|||..||||||||||||.||||||||||
Db       9338 GTGATCAAGGTGGTGCAGGGGATGCGGCATGGGTCGTTGCCGCGGACGCTGCATGTGGAT 9397

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              |||||||.||.|||||||.|||.|..||||.||||||..|||||||||.|||.|...|
Db       9398 GCGCCGTCGTCGAAGGTGGAGTGGGCTTCGGGTGCGGTGGAGCTGCTGACCGAGACCCGG 9457

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              |||||||.||...||..|.||||.||||||||||.|..||||||.||||||.|
Db       9458 TCGTGGCCGCGG---CGGGTGGAGCGGGTGCGGCGGGCCGCGGTGTCGGCGTTCGGGGTG 9514

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              ||||.||.||.||.||||||.|..|||||||.|||||||||
Db       9515 AGCGGGACCAACGCCCATGTGGTCCTGGAGGAAGCGCCGGCG 9556
```

RESULT 11
AAA09469
ID AAA09469 standard; DNA; 50937 BP.
XX
AC AAA09469;
XX
DT 29-AUG-2000 (first entry)
XX
DE Streptococcus oleandomycin gene cluster.
XX
KW oleandomycin; oleandolide; polyketide synthase; oleAI; oleAII; oleAIII;
KW PKS; type I; 8,8a-deoxyoleandolide synthase; modular; ketosynthase;
KW acyl-transferase; acyl carrier protein; inactivated; polyketide;
KW macrolactone; antibiotic; motilide; erythromycin; ss.
XX
OS Streptococcus antibioticus.
XX
FH Key Location/Qualifiers
FT CDS 152..1426
FT /*tag= a
FT /label= oleI
FT CDS complement (1528..2637)
FT /*tag= b
FT /label= oleN2
FT CDS complement (2658..4967)
FT /*tag= c
FT /label= oleR
FT CDS 5772..18224
FT /*tag= d
FT /label= ORF1
FT /product= 8,8a-deoxyoleandolide_synthase_1
FT CDS 18267..29717
FT /*tag= e
FT /label= ORF2
FT /product= 8,8a-deoxyoleandolide_synthase_2
FT CDS 29787..40346
FT /*tag= f
FT /label= ORF3
FT /product= 8,8a-deoxyoleandolide_synthase_3
FT CDS 40625..41830
FT /*tag= g
FT /label= oleP1
FT CDS 41878..43158
FT /*tag= h
FT /label= oleG1
FT CDS 43163..44443
FT /*tag= i
FT /label= oleG2
FT CDS 44433..45173
FT /*tag= j
FT /label= oleM1
FT CDS 45251..46411
FT /*tag= k
FT /label= oleY
FT CDS 46491..47714
FT /*tag= l
FT /label= oleP
FT CDS complement (47808..49517)
FT /*tag= m
FT /label= oleB
XX
PN WO200026349-A2.
XX
PD 11-MAY-2000.
XX
PF 22-OCT-1999; 99WO-US24478.
XX
PR 29-OCT-1998; 98US-0106100.
PR 16-FEB-1999; 99US-0120254.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Betlach MC, Shah SK, McDaniel R, Tang L;
XX
DR WPI; 2000-365602/31.
DR P-PSDB; AAY92707, AAY92708, AAY92709.
XX
PT Recombinant DNA compound encoding oleandolide polyketide synthase for
PT synthesizing polyketides comprising a coding sequence for a domain of a
PT loading module or any one of extender modules
XX
PS Disclosure; Page 14-26; 86pp; English.
XX
CC This is part of the Streptococcus antibioticus oleandomycin gene cluster.
CC The oleandolide polyketide synthase (PKS), also known as
CC 8,8a-deoxyoleandolide synthase, is encoded by three open reading frames
CC (ORF), designated oleAI, oleAII and oleAIII. The PKS is a type I
CC "modular" enzyme, where each ORF encodes 2 extender modules and
CC the first ORF also encodes the loading module. Each module is composed
CC of at least a ketosynthase (KS), acyl-transferase (AT) and an
CC acyl carrier protein (ACP) domain. The oleandolide PKS loading module
CC contains an inactivated KS, called KS-Q, where Q is the abbreviation for
CC glutamine, present instead of the active site cysteine required for
CC activity. The large multifunctional PKS enzymes catalyze the biosynthesis
CC of polyketide macrolactones through multistep pathways involving
CC decarboxylative condensations between acylthioesters followed by cycles
CC of varying beta-carbon processing activities. The macrolide product of
CC the PKS, 8,8a-deoxyoleandolide, is further modified by epoxidation and
CC glycosylation to yield oleandomycin, an antibacterial polyketide. The
CC invention concerns an isolated recombinant DNA compound, comprising a
CC coding sequence for a domain of loading module or any one of extender
CC modules 1-4 or 1-6, including an oleandolide PKS operably linked to a
CC promoter. Also discussed are recombinant oleandolide PKS in which the
CC module 1 KS domain is inactivated by deletion or other mutation. In

```
Db       9278 TGGCTCGGGTCGTTGAAGTCGAACATCGGGCATGCGCAGGCGGCTGCGGGTGTGGGTGGT 9337

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              || |||||| |||||  || | ||||| | |||  |||||||||||| ||||||||||
Db       9338 GTGATCAAGGTGGTGCAGGGGATGCGGCATGGGTCGTTGCCGCGGACGCTGCATGTGGAT 9397

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              ||||||| || ||||||| ||| |  |||| ||||||  |||||||| ||| |    |
Db       9398 GCGCCGTCGTCGAAGGTGGAGTGGGCTTCGGGTGCGGTGGAGCTGCTGACCGAGACCCGG 9457

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              |||||||| ||     ||  | |||| |||||||||| |  |||||| |||||||| |
Db       9458 TCGTGGCCGCGG---CGGGTGGAGCGGGTGCGGCGGGCCGCGGTGTCGGCGTTCGGGGTG 9514

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              |||| || || || |||||| |  ||||||| |||||||||
Db       9515 AGCGGGACCAACGCCCATGTGGTCCTGGAGGAAGCGCCGGCG 9556


RESULT 10
AAT80414
ID   AAT80414 standard; DNA; 44377 BP.
XX
AC   AAT80414;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Platenolide synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   platenolide synthase gene cluster; platenolide production; srmG gene;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   CDS             350..14002
FT                   /*tag=  a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in AAW22606"
FT   CDS             14046..20036
FT                   /*tag=  b
FT                   /note= "ORF2 encodes protein shown in AAW22607"
FT   CDS             20110..31284
FT                   /*tag=  c
FT                   /transl_except= (pos:20111..20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in AAW22608"
FT   CDS             31329..36071
FT                   /*tag=  d
FT                   /note= "ORF4 encodes protein shown in AAW22609"
FT   CDS             36155..41830
FT                   /*tag=  e
FT                   /note= "ORF5 encodes protein shown in AAW22610"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22606-W22610.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Example 2; Pages 110-134; 220pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. This sequence was used along with the tylG gene
CC   (see AAT80413) to create a hybrid ORF1 sequence (see AAT80415). The tylG
CC   gene is the tylactone synthase gene cluster of the invention. The tylG
CC   sequence was isolated from Streptomyces fradiae, and encodes
CC   multifunctional proteins which direct the synthesis of the polyketide
CC   tylactone. Tylactone is the basic building block of the antibiotic
CC   tylosin. The hybrid sequence can be used to transform S. ambofaciens
CC   lacking the srmG ORF1 sequence, or S. fradiae lacking the tylG ORF1
CC   sequence, so that they can produce polyketides. The DNA sequence can be
CC   modified so as to alter the type of carboxylic acids incorporated, the
CC   number of carboxylic acids incorporated and/or the post-condensation
CC   reactions performed, thereby resulting in novel tylosin-related
CC   polyketides.
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

  Query Match             54.1%;  Score 704.8;  DB 18;  Length 44377;
  Best Local Similarity   72.6%;  Pred. No. 3.3e-112;
  Matches  945;  Conservative    0;  Mismatches  342;  Indels   15;  Gaps    2;

Qy         1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
             || |||||||| ||| | |||||||| ||||||| ||||||||| ||   ||| ||| |
Db      8270 GAGCCGATCGCGATCGTGGGCATGGCGTGCCGTTACCCGGGCGGTGTGGCGTCGCCGGAC 8329

Qy        61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
             |||||||||||| |||| |||   | |||  || |  ||   |||||||| |||||||
Db      8330 GACCTGTGGGACCTGGTGGCCGGTGACGGGCACACGCTCTCCCCGTTCCCGGCCGACCGT 8389

Qy       121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
             || |||  |   ||   ||   ||| || ||| ||   |  ||| |||| |  ||||
Db      8390 GGCTGGGACGTCGAGGGGCTGTACGACCCGGAGCCGGGGGTGCCGGGCAAGAGCTATGTA 8449

Qy       181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
             | || |||||||||||| |   ||||||  | ||||||| |  |||||||| ||
Db      8450 CGGGAAGGCGGGTTCCTGCGTTCCGCGGCCGAGTTCGACGCGGAGTTCTTCGGGATATCG 8509

Qy       241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
             || || |||||   ||| ||||| |||||||||||| ||||||||||||||| ||||||
Db      8510 CCGCGCGAGGCCACGGCCATGGACCCGCAGCAGCGGTTGCTGCTGGAGACGTCGTGGGAG 8569

Qy       301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
             ||| | |||||||| || |||| ||||   ||| | |||||  |||| || || ||||||
Db      8570 GCGCTGGAGCGGGCCGGCATCGTTCCGGACTCGCTGCGCGGCACCCGGACCGGTGTCTTC 8629

Qy       361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
               |||  |  |      ||||||           |  |||  |  |  ||| | ||||
Db      8630 AGCGGCATCTCCCAGCAGGACTAC-----------GCGACCCAGCTGGGGGACGCCGCC 8677

Qy       421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
             |    |  | || ||  | ||||| || |||   ||||| ||  | || || ||  |
Db      8678 GACACCTACGGCGGGCATGTGCTCACGGGGACCCTCGGCAGTGTGATCTCCGGTCGGGTT 8737

Qy       481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
             |||||    || ||| |||| |||||||||  | || |||||||||| ||| ||| | |||
Db      8738 GCCTATGCGTTGGGGTTGGAGGGGCCGGCGCTGACGGTGGACACGGCGTGTTCGTCGTCG 8797

Qy       541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              | ||| || ||||||||||   |||||| |||||  |||||||||    || ||| ||
Db      8798 TTGGTGGCGTTGCATCTGGCGGTGCAGTCGTTGCGGCGGGGTGAGTGTGATCTGGCGTTG 8857

Qy       601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
             ||||| || ||  |||| ||| | || |      ||||| | |||||||| |||||||
Db      8858 GCCGGTGGGGTGACGGTGATGGCGACGCCGACGGTGTTCGTGGAGTTCTCGCGGCAGCGG 8917

Qy       661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
             |||||| ||| ||||||  |||||||||||| | |||  |  || ||||| || |  |||
Db      8918 GGGCTGGCGGCGGACGGGCGGTGCAAGGCGTTCGCGGAGGGTGCGGACGGGACGGCGTGG 8977
```

```
RESULT 9
AAT78508
ID   AAT78508 standard; DNA; 44377 BP.
XX
AC   AAT78508;
XX
DT   26-FEB-1998  (first entry)
XX
DE   Platenolide synthase gene cluster.
XX
KW   Platenolide synthase gene cluster; platenolide production; srmG gene;
KW   multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX
OS   Streptomyces ambofaciens.
XX
PH   Key             Location/Qualifiers
FT   CDS             350..14002
FT                   /*tag= a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in AAW23716"
FT   CDS             14046..20036
FT                   /*tag= b
FT                   /note= "ORF2 encodes protein shown in AAW23717"
FT   CDS             20110..31284
FT                   /*tag= c
FT                   /transl_except= (pos:20111..20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in AAW23718"
FT   CDS             31329..36071
FT                   /*tag= d
FT                   /note= "ORF4 encodes protein shown in AAW23719"
FT   CDS             36155..41830
FT                   /*tag= e
FT                   /note= "ORF5 encodes protein shown in AAW23720"
XX
PN   EP791656-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301066.
XX
PR   22-FEB-1996;   96US-0012050.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Burgett SG,  Kuhstoss SA,  Rao RN,  Richardson MA;
PI   Rosteck PR;
XX
DR   WPI; 1997-418047/39.
DR   P-PSDB; AAW23716-W23720.
XX
PT   DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT   for production of spiramycin-related polyketide antibiotics
XX
PS   Claim 9; Pages 8-33; 81pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. The DNA can be used to produce compounds
CC   exhibiting antibiotic activity based on the platenolide structure,
CC   including specifically the macrolide antibiotic spiramycin and spiramycin
CC   analogues and derivatives. Modifications of the platenolide synthase DNA
CC   sequence can be made so as to change the number and type of carboxylic
CC   acids incorporated into the growing polyketide chain and to change the
CC   kind of post-condensation processing that is conducted.
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

  Query Match             54.1%;  Score 704.8;  DB 18;  Length 44377;
  Best Local Similarity   72.6%;  Pred. No. 3.3e-112;
  Matches  945;  Conservative    0;  Mismatches  342;  Indels   15;  Gaps    2;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || |||||||| ||| | |||||||| ||||||| |||||||| ||   ||| ||| |
Db       8270 GAGCCGATCGCGATCGTGGGCATGGCGTGCCGTTACCCGGGCGGTGTGGCGTCGCCGGAC 8329

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              |||||||||| |||| ||| | ||| || | ||  |||||||| |||||||
Db       8330 GACCTGTGGGACCTGGTGGCCGGTGACGGGCACACGCTCTCCCCGTTCCCGGCCGACCGT 8389

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || ||| |  ||  ||  ||| || ||| ||  |  ||| |||| |  |||
Db       8390 GGCTGGGACGTCGAGGGGCTGTACGACCCGGAGCCGGGGGTGCCGGGCAAGAGCTATGTA 8449

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              | || |||||||||||| |   ||||||  | ||||||| |  ||||||| ||
Db       8450 CGGGAAGGCGGGTTCCTGCGTTCCGCGGCCGAGTTCGACGCGGAGTTCTTCGGGATATCG 8509

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              || || |||||   ||| ||||| |||||||||||| |||||||||||||||| ||||||
Db       8510 CCGCGCGAGGCCACGGCCATGGACCCGCAGCAGCGGTTGCTGCTGGAGACGTCGTGGGAG 8569

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||| | |||||||| || |||| ||||   ||| | |||||  |||| || || ||||||
Db       8570 GCGCTGGAGCGGGCCGGCATCGTTCCGGACTCGCTGCGCGGCACCCGGACCGGTGTCTTC 8629

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
                 |||  | |  |||||||                | |||  |  |  ||| | ||||
Db       8630 AGCGGCATCTCCCAGCAGGACTAC-----------GCGACCCAGCTGGGGGACGCCGCC 8677

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              |    |   | || ||  | ||||| || |||   ||||| ||  | || || ||  |
Db       8678 GACACCTACGGCGGGCATGTGCTCACGGGGACCCTCGGCAGTGTGATCTCCGGTCGGGTT 8737

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              |||||   || ||| |||| |||||||| | || |||||||||| ||| ||| | |||
Db       8738 GCCTATGCGTTGGGGTTGGAGGGGCCGGCGCTGACGGTGGACACGGCGTGTTCGTCGTCG 8797

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              | ||| || |||||||||||   |||||| |||||  |||||||||    || ||| ||
Db       8798 TTGGTGGCGTTGCATCTGGCGGTGCAGTCGTTGCGGCGGGGTGAGTGTGATCTGGCGTTG 8857

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              ||||| || ||  |||| ||| | || |       |||||  | |||||||| ||||||||
Db       8858 GCCGGTGGGGTGACGGTGATGGCGACGCCGACGGTGTTCGTGGAGTTCTCGCGGCAGCGG 8917

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              |||||| ||| ||||||  |||||||||||| | |||  |  || ||||| || |  |||
Db       8918 GGGCTGGCGGCGGACGGGCGGTGCAAGGCGTTCGCGGAGGGTGCGGACGGGACGGCGTGG 8977

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              |  ||||| || ||  |||| | || ||||||||||| | || || |||   ||  ||| |
Db       8978 GCGGAGGGTGTGGGTGTGCTGCTGGTGGAGCGGCTTTCCGACGCGCGCCGCAACGGTCAT 9037

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              |||||||||||||||| || |||||||||||| ||||||||||    |||||||||
Db       9038 CGGGTGCTGGCGGTGGTGCGGGGCAGTGCGGTCAATCAGGACGGTGCGAGCAATGGGCTG 9097

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||||||||  |||||||| ||| ||||||||||||  |||||  |||| | ||||||
Db       9098 ACGGCGCCGAGTGGTCCGGCGCAGCAGCGGGTGATCCGTGAGGCGCTGGCTGATGCGGGG 9157

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||       || || |||||||||||||||| ||||||  |||||  |||  ||||||||||
Db       9158 CTGGTGCCCGCCGACGTGGATGTGGTGGAGGCGCACGGTACGGGGACGGCGCTGGGTGAT 9217

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              ||||||||||    ||| |||| |||||||||||| |  |   |  | |   ||| ||
Db       9218 CCGATCGAGGCGGGTGCGCTGCTGGCCACGTACGGGCGGGAGCGGGTCGGCGATCCGTTG 9277

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              ||||| |||||  ||||||| |||||||||||  |   ||| || ||||||||||||||
```

```
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   platenolide synthase gene cluster; platenolide production; srmG gene;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; hybrid gene; ss.
XX
OS   Streptomyces ambofaciens.
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             350..13987
FT                   /*tag=  a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes hybrid protein shown in AAW22611"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22611.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Claim 22; Pages 178-197; 220pp; English.
XX
CC   This sequence represents a hybrid gene of the invention. This sequence
CC   was created by replacing a EcoRI-ApaI fragment of srmG ORF1 with a
CC   EcoRI-SfuI fragment from tylG ORF1. The position of the nucleotides from
CC   each of the two genes is not given in the specification. The srmG gene
CC   (see AAT80414) was isolated from Streptomyces ambofaciens, and encodes
CC   the multi-functional proteins which direct the synthesis of the
CC   polyketide platenolide. Platenolide is the basic building block of the
CC   macrolide antibiotic spiramycin. The tylG gene (see AAT80413) is the
CC   tylactone synthase gene cluster of the invention. The tylG sequence was
CC   isolated from Streptomyces fradiae, and encodes multifunctional proteins
CC   which direct the synthesis of the polyketide tylactone. Tylactone is the
CC   basic building block of the antibiotic tylosin. The hybrid sequence can
CC   be used to transform S. ambofaciens lacking the srmG ORF1 sequence, or S.
CC   fradiae lacking the tylG ORF1 sequence, so that they can produce
CC   polyketides. The DNA sequence can be modified so as to alter the type of
CC   carboxylic acids incorporated, the number of carboxylic acids
CC   incorporated and/or the post-condensation reactions performed, thereby
CC   resulting in novel tylosin-related polyketides.
XX
SQ   Sequence 13987 BP; 1556 A; 4401 C; 5727 G; 2303 T; 0 other;

  Query Match             54.1%;  Score 704.8;  DB 18;  Length 13987;
  Best Local Similarity   72.6%;  Pred. No. 3.6e-112;
  Matches  945;  Conservative    0;  Mismatches  342;  Indels   15;  Gaps    2;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || |||||||| ||| | |||||||| ||||||| |||||||| ||   ||| ||| |
Db       8255 GAGCCGATCGCGATCGTGGGCATGGCGTGCCGTTACCCGGGCGGTGTGGCGTCGCCGGAC 8314

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              |||||||||| |||| |||  | |||  || |  ||   ||||||||  |||||||
Db       8315 GACCTGTGGGACCTGGTGGCCGGTGACGGGCACACGCTCTCCCCGTTCCCGGCCGACCGT 8374

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||  |   ||   ||  ||| || ||| ||   |  ||| |||| |  ||||
Db       8375 GGCTGGGACGTCGAGGGGCTGTACGACCCGGAGCCGGGGGTGCCGGGCAAGAGCTATGTA 8434

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              | || |||||||||||| |    ||||||  | |||||||| |  |||||||| ||

Db       8435 CGGGAAGGCGGGTTCCTGCGTTCCGCGGCCGAGTTCGACGCGGAGTTCTTCGGGATATCG 8494

Qy         241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
               || || |||||   ||| ||||| |||||||||||| |||||||||||||||| ||||||
Db       8495 CCGCGCGAGGCCACGGCCATGGACCCGCAGCAGCGGTTGCTGCTGGAGACGTCGTGGGAG 8554

Qy         301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
               ||| | |||||||| || |||| ||||   ||| | |||||  |||| || || ||||||
Db       8555 GCGCTGGAGCGGGCCGGCATCGTTCCGGACTCGCTGCGCGGCACCCGGACCGGTGTCTTC 8614

Qy         361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
                  |||  |  |      ||||||                |  |||  |  |  ||| | ||||
Db       8615 AGCGGCATCTCCCAGCAGGACTAC-----------GCGACCCAGCTGGGGGACGCCGCC 8662

Qy         421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
               |    |  |  || ||  | ||||| || |||   ||||| ||  | || || ||  |
Db       8663 GACACCTACGGCGGGCATGTGCTCACGGGGACCCTCGGCAGTGTGATCTCCGGTCGGGTT 8722

Qy         481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
               |||||    || ||| |||| |||||||||| | || ||||||||||  ||| ||| | |||
Db       8723 GCCTATGCGTTGGGGTTGGAGGGGCCGGCGCTGACGGTGGACACGGCGTGTTCGTCGTCG 8782

Qy         541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
               | ||| || ||||||||||   |||||| |||||  |||||||||     || ||| ||
Db       8783 TTGGTGGCGTTGCATCTGGCGGTGCAGTCGTTGCGGCGGGGTGAGTGTGATCTGGCGTTG 8842

Qy         601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
               ||||| || ||  |||| ||| | || |      |||||  | |||||||| ||||||||
Db       8843 GCCGGTGGGGTGACGGTGATGGCGACGCCGACGGTGTTCGTGGAGTTCTCGCGGCAGCGG 8902

Qy         661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
               |||||| ||| ||||||  |||||||||||| | |||   |  || ||||| || |  |||
Db       8903 GGGCTGGCGGCGGACGGGCGGTGCAAGGCGTTCGCGGAGGGTGCGGACGGGACGGCGTGG 8962

Qy         721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
               |  ||||| || ||  ||||| |||||||||| | || || |||    ||     || |||
Db       8963 GCGGAGGGTGTGGGTGTGCTGCTGGTGGAGCGGCTTTCCGACGCGCGCCGCAACGGTCAT 9022

Qy         781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
               |||||||||||||||| || |||||||||||||| ||||||||||   |||||||||
Db       9023 CGGGTGCTGGCGGTGGTGCGGGGCAGTGCGGTCAATCAGGACGGTGCGAGCAATGGGCTG 9082

Qy         841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
               |||||||||| |||||||| ||| |||||||||||||  ||||| |||| | ||||||
Db       9083 ACGGCGCCGAGTGGTCCGGCGCAGCAGCGGGTGATCCGTGAGGCGCTGGCTGATGCGGGG 9142

Qy         901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
               ||      || || |||||||||||||||| |||||| ||||| ||| ||||||||||
Db       9143 CTGGTGCCCGCCGACGTGGATGTGGTGGAGGCGCACGGTACGGGGACGGCGCTGGGTGAT 9202

Qy         961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
               |||||||||||    ||| |||| |||||||||||| |  |   |  | |   ||| ||
Db       9203 CCGATCGAGGCGGGTGCGCTGCTGGCCACGTACGGGCGGGAGCGGGTCGGCGATCCGTTG 9262

Qy        1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
               ||||| |||||  ||||||| |||||||||||  |    ||| || ||||||||||||||
Db       9263 TGGCTCGGGTCGTTGAAGTCGAACATCGGGCATGCGCAGGCGGCTGCGGGTGTGGGTGGT 9322

Qy        1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
               || |||||| |||||  || | ||||| | |||  |||||||||||||| ||||||||||
Db       9323 GTGATCAAGGTGGTGCAGGGGATGCGGCATGGGTCGTTGCCGCGGACGCTGCATGTGGAT 9382

Qy        1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
                  ||||||| || ||||||| ||| |  |||| ||||||  |||||||| ||| |   |
Db       9383 GCGCCGTCGTCGAAGGTGGAGTGGGCTTCGGGTGCGGTGGAGCTGCTGACCGAGACCCGG 9442

Qy        1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
               |||||||| ||      ||  | |||| |||||||||| |  |||||| ||||||||  |
Db       9443 TCGTGGCCGCGG---CGGGTGGAGCGGGTGCGGCGGGCCGCGGTGTCGGCGTTCGGGGTG 9499

Qy        1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
               |||| || || || |||||| |  ||||||| |||||||||
Db       9500 AGCGGGACCAACGCCCATGTGGTCCTGGAGGAAGCGCCGGCG 9541
```

```
XX
KW   Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW   anticancer; pimaricin; ss.
XX
OS   Streptomyces natalensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..20394
FT                   /*tag= a
FT                   /product= "polyketide synthase"
XX
PN   WO200077222-A1.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-EP06227.
XX
PR   14-JUN-1999;   99EP-0201893.
XX
PA   (STAM ) DSM NV.
XX
PI   Martin JF,  Aparicio JF,  Colina AJ;
XX
DR   WPI; 2001-080693/09.
DR   P-PSDB; AAB31558.
XX
PT   New polynucleotides encoding enzymes involved in the biosynthesis of
PT   pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT   the synthesis of new compounds -
XX
PS   Disclosure; Page 53-80; 116pp; English.
XX
CC   The present sequence encodes a polyketide synthase which is associated
CC   with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC   is useful for the oxidative modification of a methyl group of a suitable
CC   compound, e.g. a bioactive compound including a secondary metabolite,
CC   antibiotics and anticancer agents. Recombinant cells comprising the
CC   gene are useful for the production of pimaricin. The polyketide synthase
CC   polynucleotide may be over expressed in Streptomyces, leading to an
CC   increase in the biosynthesis of pimaricin, as a source of primers for
CC   amplification reaction and as probes.
XX
SQ   Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;

  Query Match             54.5%;  Score 710;  DB 22;  Length 20394;
  Best Local Similarity   73.7%;  Pred. No. 4.5e-113;
  Matches  956;  Conservative    0;  Mismatches  315;  Indels   27;  Gaps    3;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              |||||| |||  ||| ||||||||   |||||  | |||||||| ||| | || ||  ||
Db       9895 GACCCGGTCGTGATCGTCGGCATGAGCTGCCGCTACCCGGGCGGCGTCGGCTCGCCCGAG 9954

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              |||||||||   ||||   |  |||||  |||||  |       ||||||||||||||
Db       9955 GACCTGTGGCGCCTGGTGTCGGAGGGCGTGGACGCGGTGTCCGACTTCCCCACCGACCGT 10014

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              ||||||  |  |||    |  ||  ||  ||||||  |  ||  |||||| |  ||  |
Db      10015 GGATGGGACGTGGAGAGCCTCTACAGCCCCGACCCCGAGGCGCTCGGCACCTCGTACACC 10074

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              |     ||  ||  ||||| |||||  ||||||  |  ||||||  |||  ||||||||  ||  ||
Db      10075 CGCTCCGGTGGATTCCTCCACGAGGCGGCGGAGTTCGACCCCGATTTCTTCGGGATGAGC 10134

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              ||  ||  |||||  |||||||   ||   | ||||||||||||||||||||||||  ||||||||
Db      10135 CCGCGCGAGGCGCTGGCGACCGACGCCCAGCAGCGGCTGCTGCTGGAGACGACCTGGGAG 10194

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||  |||||||  ||||  |||||  |||  ||||  |  ||  ||   ||||||||||||||||
Db      10195 GCCATCGAGCGCACGGGCATCGACCCGGCGTCGCTGCGGGGCAGCCGTACGGGCGTCTTC 10254

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              |||||||   |  |  ||||||||||   |   | | |  ||    ||
Db      10255 GCGGGCGTGATGTACACCGACTACGGCGACCTCCTCGTCGGCGACCAGT---------- 10303

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              |  ||||||  ||  |  ||  |||  |  |  |||  ||||  ||    ||  |||||  ||
Db      10304 -------TCGAGGGCTACCGCAGCAACGGCAGCGCGGCCAGCATCGCCTCCGGCCGGGTC 10356

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              | ||||  ||||||  |  ||  ||  ||||||  ||||  ||||||||||  ||||||  | ||
Db      10357 TCGTACACCTTCGGTTTCGAGGGTCCGGCGGTCACGGTGGACACGGCATGCTCGTCGTCC 10416

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              ||  ||  |  |||||   ||||   ||||||||||  |||||  ||||||  ||||||||  |
Db      10417 CTGGTCGCCCTGCACTGGGCGGCGCAGTCGCTGCGCTCGGGCGAGTGCTCGCTCGCGGTC 10476

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              ||  |||||  ||  ||||  ||||||||  |       |  ||||  ||||||||||  |||||  |||
Db      10477 GCGGGCGGTGTGACGGTGATGTCCACACCGACGACGTTCGTCGAGTTCTCGCGGCAACGC 10536

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              ||  |||||||  |||||||  |  ||||||||||  |  |  |||  ||||||||||  |||||||
Db      10537 GGACTGTCGGCGGACGGCCGCTGCAAGGCGTTCGCCGATGCGGCCGACGGCGTCGGCTGG 10596

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              |||||||||||||  |||||  |  |||||||  |||||||  |||    ||    |||||
Db      10597 GGCGAGGGCGTCGGCATGCTCGTACTGGAGCGTCTGTCGGACGCGCGCCGCAACGGGCAC 10656

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              ||||||||  ||||||||  ||||||||||||||  ||||||||||||||||||  |||||  |||
Db      10657 CGGGTGCTCGCGGTGGTGCGCGGCAGTGCGGTGAACCAGGACGGTGCGTCCAATGGTCTG 10716

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||||||||||||  ||  ||  |||  ||||||||||||||||||||  ||||||  ||||||
Db      10717 ACGGCGCCGAACGGCCCCGCCCAGCAGCGGGTGATCCGGCAGGCGCTGGCGAGTGCGGGG 10776

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||||  |  ||||||||||||  |  ||||||||  ||||||  |||||  ||||||||||  |||
Db      10777 CTGTCGGCGGCGGATGTGGACGCGGTGGAGGCGCACGGTACGGGTACGACGCTGGGCGAT 10836

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGG---GCAGCGGGCCGGTGACAGGCCG 1017
              ||||||||||| ||||||  ||||||||||||||  ||    |  ||||  |  |  |||  |||||
Db      10837 CCGATCGAGGCCCAGGCGCTGCTCGCCACGTATGGCCAGGAGCGACCTGAGGACCGGCCG 10896

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077
              ||  |||||||||  |  ||  ||||||||||  ||  |    |||  ||  |||||||||
Db      10897 TTGCTGCTGGGGTCGGTCAAATCCAACATCGGTCATGCGCAGGCGGCTTCGGGTGTGGCG 10956

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137
              ||  ||||||||||||||  |||||  |||||  |  ||  |||  ||||  ||||||  ||||||||
Db      10957 GGTGTCATCAAGATGGTGCTGGCGATGCGGCACGGTGTGCTGCCTCGGACGCTGCATGTG 11016

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197
              |||  |  ||||||  ||||  ||  ||||||    ||  ||  ||  ||  |||||||||     | |
Db      11017 GATGAACCGTCGTCGCATGTCGACTGGAGTGCCGGTGCCGTCGAGCTGCTGACCTCCGAG 11076

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257
              |    |||||||  ||  ||  ||       ||  |||  ||  |||||  ||  ||  ||  |||||
Db      11077 GCCGAGTGGCCGCAGGGCGAGG------GGCCGCGCCGCGCGGGCGTCTCCTCCTTCGGC 11130

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGC 1295
              ||  ||||  |||||  ||||||||||  |||||  ||  |
Db      11131 GTCAGCGGGACGAACGCGCATGTGATCCTGGAGCAGCC 11168


RESULT 8
AAT80415
ID   AAT80415 standard; DNA; 13987 BP.
XX
AC   AAT80415;
XX
DT   02-MAR-1998  (first entry)
XX
DE   Hybrid srmG/tylG ORF1.
```

```
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag=  a
FT                   /product= "AAG65266"
FT   CDS             14824..31422
FT                   /*tag=  b
FT                   /product= "AAG65267"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65266, AAG65267.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Disclosure; Page 103-149; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;

  Query Match             72.6%;  Score 944.8;  DB 22;  Length 31422;
  Best Local Similarity   83.6%;  Pred. No. 3.1e-153;
  Matches 1088;  Conservative    0;  Mismatches  202;  Indels   12;  Gaps    1;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || ||||| || ||  | || ||||| || ||||| ||||| || ||  | || ||| ||
Db      14935 GAGCCGATTGCGATTGTGGGGATGGCGTGTCGTTTTCCGGGTGGGGTGTGTTCGCCGGAG 14994

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              ||  ||||||||||||  || ||||| || || || || ||     || ||  |||  ||
Db      14995 GAGTTGTGGGAGCTGGTTGCGTCGGGTGGGGATGCGATTGGTGAATTTCCGGCCGGTCGG 15054

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||     |||  |    |   ||||||    | | | || ||||  ||  ||
Db      15055 GGGTGGGATCTGGAGGGGTTGTTTGATTCGGACCCTGACCGGTCGGGGACGTCGTACGCG 15114

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              | |   ||||||||  |  || |||| | | ||||| || | |||||||| ||||||
Db      15115 CGGTATGGCGGGTTTTTGTATGAGGCGGGGGAGTTCGATGCGGACTTCTTCGGGATCAGT 15174

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              || ||||||||  |||||||||||||||||||||||| || |||||||||||| ||||||
Db      15175 CCGCGTGAGGCGTTGGCGATGGATCCGCAGCAGCGGTTGTTGCTGGAGACGTCGTGGGAG 15234

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              |||||||||||||||| |||||||||||||| | || || |||||||||| ||||||
Db      15235 GCGTTCGAGCGGGCGGGTATCGATCCGCTGTCGATGCGTGGCTCCCGTACGGGTGTCTTC 15294

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              || || |   | | |  ||||||||  | ||  |||  |||              | ||
Db      15295 GCCGGGGTGATGTACCACGACTACGGATCCCGCCTGGGTACCAT-----------CCCC 15342

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              | ||   | |||||| |||||  || |||||  |   |||  ||||  |||||||||  ||
Db      15343 GAGGGATTCGAGGGCTACATCGGCAACGGTAGCGGCGGCGCCGTCGCGTCGGGCCGCGTC 15402

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              |||||||   |||| || || || || ||  || | ||||||||||  || ||| | |||
Db      15403 GCCTACACGCTCGGTCTCGAGGGCCCTGCCGTCTCGGTGGACACGGCATGTTCGTCGTCG 15462

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              | ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15463 TTGGTGGCGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 15522

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              |||||||| ||  |||| ||||| ||||     | | ||| ||||||||| || ||||||
Db      15523 GCCGGCGGTGTGACGGTGATGTCGACCCCGCACCTCTTCGTCGAGTTCTCACGCCAGCGC 15582

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              || |||||||||||||| | |||||| | | ||| || ||||||||||||||||  |
Db      15583 GGACTGTCGGTGGACGGCCGCTGCAAGTCCTTCGCGGGTGGAGCCGACGGCACCGGCATG 15642

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15643 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 15702

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              |||||||||||||| | |||||||||||||||| ||||||||||||||||||||||| ||
Db      15703 CGGGTGCTGGCGGTGCTGCGCGGCAGTGCGGTCAATCAGGACGGTGCGTCGAATGGGTTG 15762

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      15763 ACGGCGCCGAATGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 15822

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      15823 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCATGGGACGGGCACGACGCTGGGTGAT 15882

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              |||||||||||  |||||||||||||||||||||||||||||||||| ||||||||||
Db      15883 CCGATCGAGGCGCAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTAACAGGCCGCTG 15942

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              |||||||| ||  ||||||| |||||||| ||   |||||||||||||||||||||||
Db      15943 TGGCTGGGATCGGTGAAGTCGAACATCGGCCATGCGCAGGCTGCCGCGGGTGTGGGTGGG 16002

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16003 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 16062

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16063 GAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 16122

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      16123 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGGTC 16182

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              | ||||||||||||||||||||||||||||||||||||||
Db      16183 AGTGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 16224


RESULT 7
AAF24892
ID   AAF24892 standard; DNA; 20394 BP.
XX
AC   AAF24892;
XX
DT   20-APR-2001  (first entry)
XX
DE   Pimaricin biosynthesis associated polyketide synthase gene.
```

```
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             14824..31422
FT                   /*tag=  b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23751, AAB23752.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use -
XX
PS   Claim 2; Page 134-203; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;

  Query Match             72.6%;  Score 944.8;  DB 21;  Length 31422;
  Best Local Similarity   83.6%;  Pred. No. 3.1e-153;
  Matches 1088;  Conservative    0;  Mismatches  202;  Indels   12;  Gaps    1;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || |||||  || ||   | || |||||  || |||||  |||||  || ||   | ||  ||| ||
Db      14935 GAGCCGATTGCGATTGTGGGGATGGCGTGTCGTTTTCCGGGTGGGGTGTGTTCGCCGGAG 14994

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              ||  ||||||||||||   || |||||  || ||  ||  ||  ||     ||  ||   |||  ||
Db      14995 GAGTTGTGGGAGCTGGTTGCGTCGGGTGGGGATGCGATTGGTGAATTTCCGGCCGGTCGG 15054

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||      |||   |       |    ||||||    | | | || ||||  ||  ||
Db      15055 GGGTGGGATCTGGAGGGGTTGTTTGATTCGGACCCTGACCGGTCGGGGACGTCGTACGCG 15114

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              | |   ||||||||  |  | || |||| | | |||||  || | |||||||| ||||||
Db      15115 CGGTATGGCGGGTTTTTGTATGAGGCGGGGGAGTTCGATGCGGACTTCTTCGGGATCAGT 15174

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              || ||||||||  |||||||||||||||||||||||||  || ||||||||||||  ||||||
Db      15175 CCGCGTGAGGCGTTGGCGATGGATCCGCAGCAGCGGTTGTTGCTGGAGACGTCGTGGGAG 15234

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||||||||||||||||  ||||||||||||||||  | ||  ||  ||||||||||| ||||||
Db      15235 GCGTTCGAGCGGGCGGGTATCGATCCGCTGTCGATGCGTGGCTCCCGTACGGGTGTCTTC 15294

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              || || |   | | |  |||||||||  | ||  |||  |||                | ||
Db      15295 GCCGGGGTGATGTACCACGACTACGGATCCCGCCTGGGTACCAT------------CCCC 15342

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
```

```
                | ||   | |||||| |||||  || |||||  |   |||  ||||  |||||||||  ||
Db      15343 GAGGGATTCGAGGGCTACATCGGCAACGGTAGCGGCGGCGCCGTCGCGTCGGGCCGCGTC 15402

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              |||||||   |||| || || || ||  ||  || | ||||||||||  || ||| | |||
Db      15403 GCCTACACGCTCGGTCTCGAGGGCCCTGCCGTCTCGGTGGACACGGCATGTTCGTCGTCG 15462

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              | ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15463 TTGGTGGCGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 15522

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              |||||||| ||  |||| |||||  |||     | | |||  |||||||||| || ||||||
Db      15523 GCCGGCGGTGTGACGGTGATGTCGACCCCGCACCTCTTCGTCGAGTTCTCACGCCAGCGC 15582

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              || |||||||||||||| | |||||| | | | ||| || ||||||||||||||||  |
Db      15583 GGACTGTCGGTGGACGGCCGCTGCAAGTCCTTCGCGGGTGGAGCCGACGGCACCGGCATG 15642

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15643 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 15702

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              |||||||||||||| | |||||||||||||| ||||||||||||||||||||||||  ||
Db      15703 CGGGTGCTGGCGGTGCTGCGCGGCAGTGCGGTCAATCAGGACGGTGCGTCGAATGGGTTG 15762

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15763 ACGGCGCCGAATGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 15822

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      15823 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCATGGGACGGGCACGACGCTGGGTGAT 15882

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              |||||||||| ||||||||||||||||||||||||||||||||||| ||||||||||||||
Db      15883 CCGATCGAGGCGCAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTAACAGGCCGCTG 15942

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              |||||||| ||  |||||||| |||||||| ||  |   ||||||||||||||||||||||
Db      15943 TGGCTGGGATCGGTGAAGTCGAACATCGGCCATGCGCAGGCTGCCGCGGGTGTGGGTGGG 16002

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16003 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 16062

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16063 GAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 16122

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16123 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGGTC 16182

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              | ||||||||||||||||||||||||||||||||||||||||
Db      16183 AGTGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 16224
```

```
RESULT 6
AAH79278
ID   AAH79278 standard; DNA; 31422 BP.
XX
AC   AAH79278;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 2.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
```

```
XX
PR   06-JUL-1998;   98GB-0014622.
XX
PA   (BIOT-) BIOTICA TECHNOLOGY LTD.
PA   (PFIZ ) PFIZER INC.
XX
PI   Kellenberger JL,  Leadlay PF,  Staunton J,  Stutzman-Engwall KJ;
PI   McArthur HAI;
XX
DR   WPI; 2000-182117/16.
XX
PT   Mutated Type I polyketide synthase containing a polylinker site in an
PT   extension module for replacement of a reductive loop sequence, for
PT   producing polyketides, e.g. B1 avermectin -
XX
PS   Disclosure; Fig 7a-f; 75pp; English.
XX
CC   The present sequence is that of DNA encoding the first 2 modules
CC   of the avermectin polyketide synthase (PKS) of Streptomyces
CC   avermitilis.  The invention relates to nucleic acids encoding a
CC   Type I PKS such as avermectin in which a polylinker with multiple
CC   restriction sites replaces or 1 more PKS genes encoding enzymes
CC   associated with reduction.  Novel PKS are provided in which in
CC   which the reductive loop in a selected module of the Type I PKS is
CC   replaced with the equivalent segment from the same or different
CC   PKS gene cluster or by a mutated or synthetic segment.  Vectors and
CC   host cells, and methods for producing novel polyketides by
CC   culturing host cells are claimed.  The polyketides obtained are
CC   useful as antibiotics and insecticides.  Fermentation products
CC   containing C22-C23 dihydroavermectin, ivermectin and B1
CC   avermectins are claimed.
XX
SQ   Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;

  Query Match             99.8%;  Score 1298.8;  DB 21;  Length 12381;
  Best Local Similarity   99.8%;  Pred. No. 9.3e-214;
  Matches 1300;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2258 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 2317

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       2318 GACCTGTGGGAGCTGCCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 2377

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2378 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 2437

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2438 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 2497

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2498 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 2557

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2558 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 2617

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2618 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 2677

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2678 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 2737

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2738 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 2797

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2798 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 2857

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2858 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 2917

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2918 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 2977

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2978 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 3037

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3038 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 3097

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3098 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 3157

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3158 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 3217

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3218 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 3277

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3278 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 3337

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3338 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 3397

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3398 GAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 3457

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3458 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 3517

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              |||||||||||||||||||||||||||||||||||||||||||
Db       3518 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 3559


RESULT 5
AAA92302
ID   AAA92302 standard; DNA; 31422 BP.
XX
AC   AAA92302;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag=  a
```

```
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65268.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 10; Page 149-167; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is modified version of a
CC   fragment of the S. avermitilis genome.
XX
SQ   Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;

  Query Match             99.9%;  Score 1300.4;  DB 22;  Length 11916;
  Best Local Similarity   99.9%;  Pred. No. 4.9e-214;
  Matches 1301;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 1500

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 1560

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 1620

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 1680

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 1740

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 1800

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 1860

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 1920

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       1921 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGGGCTCGGCATCG 1980

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 2040

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 2100

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 2160

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 2220

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 2280

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 2340

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 2400

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 2460

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 2520

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 2580

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 2640

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 2700

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              |||||||||||||||||||||||||||||||||||||||||||
Db       2701 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 2742


RESULT 4
AAZ58381
ID   AAZ58381 standard; DNA; 12381 BP.
XX
AC   AAZ58381;
XX
DT   23-MAY-2000  (first entry)
XX
DE   Streptomyces avermitilis avermectin polyketide synthase modules 1+2.
XX
KW   Polyketide synthase; avermectin; insecticide; ss.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200001827-A2.
XX
PD   13-JAN-2000.
XX
PF   06-JUL-1999;   99WO-GB02158.
```

```
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65264, AAG65265.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Example 2; Page 58-123; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;

  Query Match             100.0%;  Score 1302;  DB 22;  Length 30690;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-214;
  Matches 1302;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 1500

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 1560

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 1620

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 1680

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 1740

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 1800

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 1860

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 1920

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 1980

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 2040

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 2100

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 2160

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 2220

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 2280

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 2340

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 2400

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 2460

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 2520

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 2580

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 2640

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 2700

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              |||||||||||||||||||||||||||||||||||||||||||
Db       2701 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 2742


RESULT 3
AAH79279
ID   AAH79279 standard; DNA; 11916 BP.
XX
AC   AAH79279;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequence derivative SEQ ID NO: 3.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11916
FT                   /*tag=  a
FT                   /product= "AAG65268"
FT                   /partial
XX
```

```
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23749, AAB23750.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use -
XX
PS   Claim 2; Page 66-134; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;

  Query Match           100.0%;  Score 1302;  DB 21;  Length 30690;
  Best Local Similarity  100.0%;  Pred. No. 2.5e-214;
  Matches 1302;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 1500

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 1560

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 1620

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 1680

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 1740

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 1800

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 1860

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 1920

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 1980

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 2040

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       2041 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 2100

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 2160

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 2220

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 2280

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 2340

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 2400

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 2460

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 2520

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 2580

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 2640

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 2700

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              |||||||||||||||||||||||||||||||||||||||||||
Db       2701 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 2742


RESULT 2
AAH79277
ID   AAH79277 standard; DNA; 30690 BP.
XX
AC   AAH79277;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 1.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag=  a
FT                   /product= "AAG65264"
FT                   /partial
FT   CDS             11971..30690
FT                   /*tag=  b
FT                   /product= "AAG65265"
XX
PN   WO200162939-A1.
XX
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 259.003 Seconds
(without alignments)
11320.736 Million cell updates/sec

Title: US-09-914-286-1_COPY_1441_2742
Perfect score: 1302
Sequence: 1 gacccgatcgccatcatcgg.....ttttggaggaggcgccggcg 1302

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1302	100.0	30690	21	AAA92301	S. avermitilis ave
2	1302	100.0	30690	22	AAH79277	Streptomyces averm
3	1300.4	99.9	11916	22	AAH79279	Streptomyces averm
4	1298.8	99.8	12381	21	AAZ58381	Streptomyces averm
5	944.8	72.6	31422	21	AAA92302	S. avermitilis ave
6	944.8	72.6	31422	22	AAH79278	Streptomyces averm
7	710	54.5	20394	22	AAF24892	Pimaricin biosynth
8	704.8	54.1	13987	18	AAT80415	Hybrid srmG/tylG O
9	704.8	54.1	44377	18	AAT78508	Platenolide syntha
10	704.8	54.1	44377	18	AAT80414	Platenolide syntha
11	689	52.9	50937	21	AAA09469	Streptococcus olea
12	676.4	52.0	43280	18	AAT80413	Tylactone synthase
13	662.6	50.9	5544	22	AAF24891	Pimaricin biosynth
14	655	50.3	53789	19	AAV21187	Amycolatopsis medi
15	650.4	50.0	14775	22	AAF88338	S. spinosa DNA fra
16	650.4	50.0	50000	22	AAF88313	S. spinosa DNA fra
17	650.4	50.0	50000	22	AAF88316	S. spinosa DNA fra
18	650.4	50.0	80161	20	AAZ21501	DNA fragment of Sa
19	649	49.8	16767	22	AAF88339	S. spinosa DNA fra
20	646.4	49.6	9513	22	AAF88337	S. spinosa DNA fra
21	645.4	49.6	6459	22	AAF88336	S. spinosa DNA fra
22	642.2	49.3	38506	21	AAA75633	Nucleotide sequenc
23	642.2	49.3	38506	21	AAZ56001	Recombinant cosmid
24	641.6	49.3	15872	21	AAZ87283	S. venezuelae vep
25	639	49.1	36778	21	AAZ87318	S. venezuelae pik
26	639	49.1	37948	21	AAZ87285	S. venezuelae pik
27	638.2	49.0	27541	22	AAD17185	Streptomyces nours
28	638.2	49.0	125401	22	AAD17186	Streptomyces nours
29	637	48.9	13842	21	AAZ87297	S. venezuelae macr
30	636	48.8	65140	22	AAD17184	Streptomyces nours
31	634.2	48.7	11220	21	AAZ87298	S. venezuelae macr
32	633.6	48.7	15872	18	AAT68715	Streptomyces venez
33	625.4	48.0	4689	21	AAZ87299	S. venezuelae macr
c 34	623	47.8	77536	21	AAA14651	Nucleotide sequenc
35	621.4	47.7	4403765	22	AAI99683	Mycobacterium tube
36	621.4	47.7	4411529	22	AAI99682	Mycobacterium tube
37	618.4	47.5	4571	21	AAA14662	Nucleotide sequenc
38	616.8	47.4	4478	21	AAA14661	Nucleotide sequenc
39	615.6	47.3	4466	21	AAA14663	Nucleotide sequenc
40	615.6	47.3	4547	21	AAA14664	Nucleotide sequenc
41	615.6	47.3	77536	21	AAA14651	Nucleotide sequenc
42	609.8	46.8	7788	22	AAF88335	S. spinosa DNA fra
43	609.8	46.8	50000	22	AAF88312	S. spinosa DNA fra
44	608.2	46.7	29879	14	AAQ46806	eryA region of S.
45	604.8	46.5	3978	21	AAC55785	Type I polyketide

ALIGNMENTS

RESULT 1
AAA92301
ID AAA92301 standard; DNA; 30690 BP.
XX
AC AAA92301;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical; ds.
XX
OS Streptomyces avermitilis.
XX
FH Key Location/Qualifiers
FT CDS 1..11919
FT /*tag= a
FT /note= "avermectin aglycon synthase protein"
FT CDS 11971..30690
FT /*tag= b
FT /note= "avermectin aglycon synthase protein"
XX
PN WO200050605-A1.
XX
PD 31-AUG-2000.
XX
PF 23-FEB-2000; 2000WO-JP01041.
XX
PR 24-FEB-1999; 99JP-0046961.
XX

```
Db      94523 CTGTGGCTGGGGTCGGTGAAGTCCAACCTGGGCCACACGCAGGCGGCGGCCGGGGCCGCG 94464

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137
Db      94463 GGCATCATCAAGATGATCATGGCTATGCGGTACGGGGTGCTGCCGAGGACGCTGCATGTG 94404

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197
Db      94403 GACCGGCCGTCGCCCGAGGTGGACTGGTCCCCGGGCACCGTGGAACTGCTCACCGAGGAG 94344

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257
Db      94343 CGGGAGTGGCCG------GACGCGGGCCGTCCGCGCCGGGCCGCCGTGTCCTCGTTCGGG 94290

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGC 1301
Db      94289 ATCAGCGGCACCAACGCCCATGTCATCCTCGAACAGCCGCCTGC 94246


RESULT 15
AX250262
LOCUS       AX250262               30000 bp    DNA     linear   PAT 05-OCT-2001
DEFINITION  Sequence 2 from Patent WO0168867.
ACCESSION   AX250262
VERSION     AX250262.1  GI:15984064
KEYWORDS    .
SOURCE      Streptomyces cinnamonensis.
  ORGANISM  Streptomyces cinnamonensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 30000)
  AUTHORS   Leadlay,P.F., Staunton,J. and Oliynyk,M.C.
  TITLE     Polyketides and their synthesis
  JOURNAL   Patent: WO 0168867-A 2 20-SEP-2001;
            Biotica Technology Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .30000
                     /organism="Streptomyces cinnamonensis"
                     /db_xref="taxon:1900"
                     /note="Nucleotides 30001 - 60000 of the monensin
                     biosynthetic gene cluster"
BASE COUNT     3911 a  10993 c  10862 g   4234 t
ORIGIN

  Query Match           52.0%;  Score 677;  DB 6;  Length 30000;
  Best Local Similarity   71.8%;  Pred. No. 1.1e-76;
  Matches  935;  Conservative    0;  Mismatches  350;  Indels   18;  Gaps    3;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
Db      17944 GACCCGGTCGCCATCGTCGCCATGACGTGCCGGTTCCCGGGCGGCGTCGCGTCCCCGGAC 18003

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
Db      18004 GACCTGTGGGACCTGCTCGACGCACGCAAGGACGCGATGGGCGCCTTCCCCACCGACCGC 18063

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
Db      18064 GGCTGGGACCTGGAACGCCTCTTCCACCCCGACCCGGACCACCCCGGCACCAGCTACACC 18123

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
Db      18124 GACCAGGGCGGATTTCTTCCCGACGCGGGTGATTTCGATGCGGCGTTCTTCGGGATCAAT 18183

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
Db      18184 CCGCGGGAGGCGCTGGCGATGGATCCGCAGCAGCGGTTGTTGCTGGAGGCGTCGTGGGAG 18243

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
Db      18244 GTGTTGGAGCGTGCGGGTATCGATCCGACGACGCTCAAGGGCACCCCGACCGGCACCTAC 18303

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
Db      18304 GTGGGCCTCATGTACCACGACTACGCCAAGTCCTTCCCCACGGC-----------CGAC 18351

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
Db      18352 GCCCAGTTGGAGGGCTACTCCTACTTGGCGAGCACCGGCAGCATGGTCTCCGGCCGCGTC 18411

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
Db      18412 GCCTACACCCTGGGCCTTGAAGGTCCGGCGGTGACGGTCGACACCGCGTGCTCCTCCTCC 18471

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
Db      18472 CTGGTCTCCATCCACCTGGCGACGCAGGCACTCCGGCACGGCGAGTGCGACCTCGCCCTG 18531

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
Db      18532 GCAGGCGGTGTGACCGTCATGGCCGACCCGGACATGTTCGCGGGCTTCTCGCGCCAGCGC 18591

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
Db      18592 GGCCTCTCACCTGACGGCCGCTGCAAGGCCTACGCCGCCGCGGCCGACGGAGTCGGATTC 18651

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
Db      18652 TCCGAGGGAGTGGGCGTATTGCTCCTTGAGCGGTTGTCGGATGCGCGGCGTCATGGGCGT 18711

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
Db      18712 CGGGTGTTGGGTGTGGTGCGGGGTTCGGCGGTGAATCAGGACGGTGCGAGTAATGGGTTG 18771

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
Db      18772 ACGGCGCCGAATGGTCCGTCGCAGGAGCGGGTGATTCGTCAGGCGTTGGCCAGTGGTGGG 18831

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
Db      18832 TTGTCGTCGGTGGATGTTGATGTGGTGGAGGGGCATGGGACGGGGACCACGTTGGGTGAT 18891

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCA---GCGGGCCGGTGACAGGCCG 1017
Db      18892 CCGATCGAGGCGCAGGCTCTGCTGGCCACATATGGGCAGGGGCGTCCGGAGGACCGTCCG 18951

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077
Db      18952 TTGTGGTTGGGGTCGGTGAAGTCGAACATTGGTCATACGCAGGCGGCTGCGGGTGTTGCG 19011

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137
Db      19012 GGTGTCATCAAGATGGTGATGGCGATGCGGCATGGTGTGGTGCCGGCGAGTTTGCATGTG 19071

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197
Db      19072 GATGTGCCGTCGCCGCATGTGGAGTGGGATTCGGGTGCGGTGCGGTTGGCGGTTGAGTCG 19131

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257
Db      19132 GTGCCATGGCC---GCAGGTGGAGGGTCGTCCGCGTCGGGCGGGTGTGTCGTCGTTCGGC 19188

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGG 1300
Db      19189 GCTTCGGGGACGAATGCGCACGTGATCGTGGAGTCTGTTCCCG 19231


Search completed: June 17, 2003, 17:55:28
Job time : 2533.07 secs
```

```
                LSPVPSVPTVPSVSSASSAASASSAASAASAVSAAPAASAAFAKAALAWAYRDEPVLF
                DVACGFAVLGESATPARVACLVDRGADAVDQVMTALDTAGLLAGGTFRSALVGKALLE
                HVDVETRAELRRRAARLLHGDGAPATVVARHLLAAPVAEPWAGRTLLAAAEKAMQEGR
                EELSLDCLRLAGQSATTEQERAAVVMARVKIGWEIDPRLVSPWLDELVAALRRGDLRS
                GDAAWIVKHLAWQDRFPEAVESLSTLAAAGDAGEPGVEAELHFVRHWLRHSCPPLEAS
                VPGPSNPTRPRTPRLRRVSPASYAGEVLGKVLTEGPDDHTVAMAEEILRGCRYSETVV
                EALEAALLALVYAGRPDRAAFWCDGLLQQAGTRSTGTSVAILSGIRAEIALRQGALPA
                VEQYARSALSAISLPGWGVAVGSPLAALVHSTTASGMSGVAGAWLNQAVPDGMFRTRH
                GLLYTHARGHYHLAVDRPAAALDDFLTCGGLAKEWGMDVPTFLPWRTSAAQAQLALGN
                PGQARALVREQLGRPGGSTARARGVALRLLAATGEPGRRPALLRESVNLLESCGDQVE
                LLRALADRCQALDHVGAPAKARMTARHARNVADACGGDALFRKLFRDEVPDREGDPDG
                LTGEEQGSASLTAAERRVTALAALGHSNREIGRKLFITKSTVEQHLTRVYRKLGVRNR
                ADLGDLLAGTNLGAGTPAMAMASSAVAG"
CDS             10933. .11271
                /codon_start=1
                /transl_table=11
                /product="P450-like protein"
                /protein_id="BAB69191.1"
                /db_xref="GI:15823974"
                /translation="MEVGGQTIRAGEGVLCTLSTANRDEGAFTDPDELDLRRDARSHL
                SFGFGIHQCIGQSLARAEPQIALETLLRRLPGLRLAASFEELRFRDGVGFYGVKELPD
                PCPALPPDRP"
gene            11458. .29898
                /gene="olmA1"
CDS             11458. .29898
                /gene="olmA1"
                /codon_start=1
                /transl_table=11
                /product="modular polyketide synthase"
                /protein_id="BAB69192.1"
                /db_xref="GI:15823975"
                /translation="MACRLPGAAGPAEFLALLRNGEDAVGAAPGDRRYAPRRGGFLDS
                VDEFDAGFFGISPREAAAMDPQQRLVLELSWEALEDAGIPPARLGGSGAGVFVGAIAD
                DYATLSRAAGADAATPQTTTGLNRGIIANRVSYTLGLRGPSFTVDSGQSSSLVAVHLA
                AESLRRGECSVALAGGVNLILAEDSTVAVERFGALSPDGYCYTFDARANGYVRGEGGG
                VVVLKPLADAIADGDDIHCILAGSAVNNDGGGEGLTVPDRAGQEAVLTAAYEQAGISP
                AAVGYVELHGTGTPTGDPVEAAAVGAVLGAGRPAGQPLLVGSVKTNIGHLEGAAGIAG
                LLKAVLAVRHREIPATLHHASPNPAIPMAELGIGVNTALTPWPEQAGPLVVGVSSFGM
                GGTNCHVVLTEPAGYDAAARPATAGIRPAGTPAPLLVSGRDEQALRAQARKLRDHLDT
                HPGLPLRDVAHTLAAARAPLARRAVLLAADREDVSAGLARLADGTPGADLVRGTVSPG
                ALAFVFTGQGSQRPGMTAELYRTSAEYAAALDEVCASLDPWLRRPLRDVLFAAEGSDE
                AALLDRTEFTQPALFAVEVALFRYAEHCGLTPRFLLGHSVGELAAAHVAGVLSLADAC
                TLVAARGRLMQDRPATGAMVSVQATEDELAPFVSESVTLAALNGPASTVIAGDEDAVL
                AVAAHWAARGRKTKRLRVSHAFHSPHMDGMLTEFRRIAAGLTFGAPRIPIVSNETGTL
                LTEADARSPEYWVRHARAAVRFLDGVRLLEDQGVTTLLELGPDGVLSAMARDCLRDAG
                TVCVPLLRGRDRTEPEALLAALAGLYVRGVPMDWERTATEAGARRVPLPTYAFQRRRH
                WLPEPSASASRPVVNGGGHAGQAGHAGQAGRAGQAGHAGHDGEPAPAVRDRTTPELVR
                DRTTLELVRDATAQVLGHASPDDVATHEAFRELGFSSLMLAELGERLTEATGRRVPTT
                LLFDHPTPDALARRLEADGAERTDAGPVSPPAAAAHDDDPIVIVGMGCRLPGGVRSPE
                DLWRFVAAGGDAISALPVDRGWEFAADFPGEGGFLADVAEFDAGFFGISPREALAMDP
                QQRLLLETSWEALERAGVDALKLRGSRTGVFVGASPSEYGPRLHETSDADGHVLTGTA
                PSVLSGRIAYVLGLEGPALTVDTACSSSLVALHLAAQALRGGECDLALAGGVTVMATS
                GMFAEFARQGGLAPDGRCKAFADGADGTGWGEGVGMLAVQRLSDAVRDGRPVLAVVRG
                SAVNSDGASNGLTAPNGPSQQRVIRQALASAGLSAADVDMVEAHGTGTSLGDPIEAQA
                LLATYGQDRPDDRPLWLGSVKSNIGHTQAAAGVAGVIKSVLALRNGLLPKTLHVDEPT
                PEVDWSAGAVELLTEGREWPETDGPRRAGVSAFGISGTNAHVILEQAPSVEEPVTAPA
                DDGLLVPWVVSARSEEALRAQARRLADHVRESGLRPADVGLSLAVTRAGLEHRAVLVA
                EDEAGFLAHLDALAEGRTPAGVVVEGGTAFLFSGQGSQRAGMGRELYSAHPVFAAAFD
                EVCAAFDGLLPGALREVVFGGAGLDRTEWAQPALFALEVALFELVSSWGVRADVLLGH
                SIGELAAAYVAGVWSLADACRIVAARGRLMQALPEGGAMVAVEAAEDELPDLPGGVSV
                AAVNGPRSLVLSGDEEPVTALAQAFAEQGRRTKQLTVSHAFHSARMEPMLAEFAETLA
                AVEFRPPRIPVVSNVTGEMAGDEFTTPAYWVRHVREAVRFADGVGAAHAAGVRRYLEL
                GPDGTLTTLAHDTAHDAGRDVAAVPALRGDQPETRTVVTAVGRLYATGLTVDWPAFFA
                PYGARPTELPTYAFQHRRYWLEPVAPSSAPPANPLRYRTAWASVPDGPRPSLSGTWAV
                VVPASGSYEDIVAGVTAALRDHGARPVVVEVDPAGPERADVADTLRAALGEDAGGVVS
                LLGTAGSLSSGVAATAVLVQAWTDLGGGARLWCLTRGAVSVSPSDPLAAPAQAQLWGM
                GRVAALEHPERWGGLVDLPAVPDDRAWTRLCAVLAGGSGEDQVAVRAAGLFARRLGRA
                GTRADAAPAGPSPWRTDGTVLVTGGTGALGAHLAQWLAAAGAAHVLLTSRRGPDADGA
                AELTARLRETGTEVTVAACDVTDRDALAELLSGLPADRPLTGVVHAAGVLDDGVLDSL
                TPDRFAAVAGPKVTGAWHLHELTRDLDLSAFVLFSSFAGTVGLAGQANYAAANAYLDA
                LAVHRAQLGLPATAVAWGSWSGAGMAADTEAARRQLARTGLVPLAPDAALAALGQVMD
                DGETAVTVADVDWDRFAAGFGPGRPSPLLTGVPEVRRIEERETGRAGAPAETDRLGGL
```

```
Query Match           52.8%;  Score 688;  DB 1;  Length 104326;
Best Local Similarity  72.3%;  Pred. No. 3.2e-78;
Matches  943;  Conservative    0;  Mismatches  340;  Indels   21;  Gaps    3;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || |||||||| ||| | |||||||| |||||||  || ||||| ||     || ||| ||
Db      95531 GAGCCGATCGCGATCGTGGGCATGGCGTGCCGTTATCCCGGCGGGGTGGCCTCGCCGGAG 95472

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              || |||||||||||||| |||| | ||  | ||||| |||||    | |||   ||||||
Db      95471 GAGCTGTGGGAGCTGGTCGCCGCAGGGCGGGACGCGATCGGTGACATGCCCGAGGACCGC 95412

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||  |  |||  ||   ||| || ||| |   | | ||| ||||| |  ||   |
Db      95411 GGCTGGAACGTGGAGGAGTTGTACGACCCGGAGCTGGCCCGGCCGGGCACCTCGTACGTG 95352

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              | ||  ||||| |||||  |||||||||  | ||||| | |  |||||||| |||
Db      95351 CGGGAGGGCGGCTTCCTGTACGACGCGGCCGAGTTCGACCCGGAGTTCTTCGGGATCTCG 95292

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              || |||||||| |||||||||| |||||||||||||||| |||||||||| ||||||
Db      95291 CCGCGTGAGGCGCTGGCGATGGACCCGCAGCAGCGGCTGCTCCTGGAGACGTCGTGGGAG 95232

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||| | |||||||| || |||||||| | |||| |||||   ||| || || || |||
Db      95231 GCGCTGGAGCGGGCCGGCATCGATCCGCGGGCGGTGCGCGGCAGCCGGACCGGGGTGTTC 95172

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              || || |   | | |  |||||||||                ||| |  | | |     ||
Db      95171 GCCGGGGTGATGTACCACGACTACGGCG------------CCGACCAGGCCGTACTGCCC 95124

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              | ||| || |||||   ||||   |||||  |  | ||||| |||   ||||| || |||
Db      95123 GAGGAGGTCGAGGGGTTCATCGGGACCGGCTCGGCCGGCAGTGTCGCCTCGGGGCGGATC 95064

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              || | ||   | || ||||| || || ||| |||| | ||||| | ||||||| | |||
Db      95063 GCGTTCACGCTGGGTCTGGAGGGTCCCGCGGTCACGCTCGACACCGCGTGCTCGTCCTCG 95004

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              || ||| ||||||| ||||||   ||| |||||||  | |||||||||   || ||||||
Db      95003 CTGGTGGCGCTGCACCTGGCGGCGCAGGCGCTGCGCACCGGTGAGTGCGACCTGGCGCTG 94944

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              ||||| |||||  | || ||||| || |  ||  ||||| ||||||| |||||||||| ||
Db      94943 GCCGGGGGCGTGACCGTGATGTCGACGCCGGGTGTGTTCGTCGAGTTGTCCCGGCAGGGC 94884

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              ||||||||    ||||| | ||| | | | | |    || ||||| || ||||||
Db      94883 GGGCTGTCCCCCGACGGCCGCTGCCGGTCCTTCGCCGAGAGCGCGGACGGTACGGGCTGG 94824

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              |||||||| | |||| | ||| |||| |||||| ||||||| ||     ||     |||||
Db      94823 GGCGAGGGTGCCGGGCTCCTGCTGGTCGAGCGGCTGTCGGACGCCCGCCGCAACGGGCAC 94764

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              |  |||||||||| || || ||||  ||||||||||||||||| || || || ||||||
Db      94763 CCCGTGCTGGCGGTCGTCCGGGGCACCGCGGTCAACCAGGACGGCGCCTCCAACGGGCTG 94704

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              || ||||| |||| || || ||| |||||||||    ||||| |||||||||| ||
Db      94703 ACCGCGCCCAACGGGCCCGCGCAGCAGCGGGTGATCGAACAGGCGCTGGCGAACGCGCGG 94644

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              |    |  || || || || |||||||||| |||||||||| |||||||||| |||
Db      94643 CTCGGTGGCGCCGACGTCGACGTGGTGGAGGCGCACGGGACGGGGACGACGCTGGGCGAT 94584

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACG---GGCAGCGGGCCGGTGACAGGCCG 1017
              |||||||||| |||||  |||| ||||| |||||   || |||||  ||  ||   |||
Db      94583 CCGATCGAGGCGCAGGCCCTGCTGGCCACCTACGGCCGGGAGCGGAGCGCCGAGCAGCCA 94524

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077
```

```
            || || || ||||||||| |||||||| |    ||| ||| |||||  ||||||||||||
Db      14475 GGTGTGATGAAGATGGTGCTGGCGTTGGGTCGGGGTGTGGTGCCGAAGACGTTGCATGTG 14534

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197
             ||| ||||||| ||||| |||||||||| || || | ||||  | ||  |   ||||||
Db      14535 GATGAGCCGTCACCGCATGTGGACTGGTCGGCTGGTGAGGTGGAGTTGGCGGTTGAGGCG 14594

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257
             ||||||||| || |       ||  |||||| |||||||||| || ||||||||||||||||
Db      14595 GTGCCGTGGTCGCG------GGGTGGGCGGGTGCGGCGGGCTGGTGTGTCGTCGTTCGGG 14648

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              ||| ||||||||||||||||||||  |||||||||||| |||
Db      14649 ATCAGCGGCACGAATGCGCATGTGATCGTGGAGGAGGCGCCTGCG 14693


RESULT 14
AB070940/c
LOCUS       AB070940               104326 bp    DNA     linear   BCT 22-JAN-2002
DEFINITION  Streptomyces avermitilis oligomycin biosynthetic gene cluster.
ACCESSION   AB070940
VERSION     AB070940.1  GI:15823967
KEYWORDS    .
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
  TITLE     Genome sequence of an industrial microorganism Streptomyces
            avermitilis: deducing the ability of producing secondary
            metabolites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
  MEDLINE   21477403
REFERENCE   2  (bases 1 to 104326)
  AUTHORS   Ikeda,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .104326
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            complement(1. .1134)
                     /gene="mcmA"
     CDS             complement(1. .1134)
                     /gene="mcmA"
                     /note="methylmalonyl-CoA mutase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB69185.1"
                     /db_xref="GI:15823968"
                     /translation="MTGYSTAADANARFRQLVGSGTRAVPVVFDLPTRAGLDSDAPAA
                     RGLVGRRGVAVDSIDDMRVLFGGLRLDRVVPSLAVDAPAAPLLVLYQLVAEEHGAPAQ
                     RLAGAVRGDVFKEFLARGPGVFPLRPSLRLAMNVFAYCQAELPRWQCLSVCGHHLAGT
                     GADPALEVAFTVAGGIEYLRAAVTAGLDVRTVAPRLTLCLVGGTTGRENRAKLRATRR
                     VWTRTANERFGSGRTAPRLYVRTPRDGLATLACTEDIEAGATALVDRIERLGGTLTVL
                     EHGLAGQLLGRPHGDRPATPRPHPPHQDSLDLQARQTERLAKLRAWRVQPAVDEALLR
                     LRKAAQGDDNVLYPMREALAARATVGEVCDALRSTWGTGPSAT"
     gene            complement(1131. .1886)
                     /gene="pptA"
     CDS             complement(1131. .1886)
                     /gene="pptA"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphopantheteinyl transferase"
                     /protein_id="BAB69186.1"
                     /db_xref="GI:15823969"
                     /translation="MRAHEVFADVPVSSLHPREAALVQDAVESRRREFATGRHCARRA
                     LASLGVPPGPLLATPRGAPRWPAGVCGAITHCAGYRAAAVARAPGIRAVGIDAEPDRP
                     LPEGVLEAVALPAEQFQVKEHTARHPGVVWGKLLFSAKESVYKAWFPATGRPLGFEEA
                     LIDLDPVSSTFRVRVLPTALRAAEPELRDLHGRWLVRHGIIVTAVVVDGPGRDTPHAP
                     SPYDRTKHSAPTGRSADRPGGRPTDRATDRPTA"
gene            complement(1910. .2875)
                /gene="simX4 homolog"
CDS             complement(1910. .2875)
                /gene="simX4 homolog"
                /codon_start=1
                /transl_table=11
                /product="SimX4 homolog"
                /protein_id="BAB69187.1"
                /db_xref="GI:15823970"
                /translation="MSQDRHDGSTAALGGPSHGGRLLAVSDIHVRHRQNRDIVQDLHP
                GCAEDWLIVAGDVAESVADITWALSLLRDRFAKVIWAPGNHELWTAPGDPDELRGERK
                YRHLVDLCRALGVVTPEDPYPVWHGAGGPVAVAPLFLLYDYSFLPPGAWSKSHALRLA
                DDAGVICTDEFLLHPDPYPSREAWCRARVRATRARLDALPDDLPTVLVNHFPLVREPT
                RVLRHPEFALWCGTTATADWHLRYRAAAVVHGHLHIPRTIRCDDVPHHEVSLGYPREW
                QRYGRRPDWPVQILTPAPAPAPAPVPVPVPAPPGHRTASPADRRP"
gene            3213. .3971
                /gene="olmC"
CDS             3213. .3971
                /gene="olmC"
                /codon_start=1
                /transl_table=11
                /product="thioesterase"
                /protein_id="BAB69188.1"
                /db_xref="GI:15823971"
                /translation="MSANRVDGNLWVRQFHPAPTAATKLVCLPHAGGSASYFLPVSKA
                LAPGVDVLAIQYPGRQDRRHEKCIDNIAELADAVTAVLEPLADRPLVLFGHSMGATLA
                FEVALRLERKGIVPLALFASGRRAPSRHRDESVHLRGDDEIIAELKTLSGTRSEVLGD
                EEILRMVLPAIRSDYKAAETYRYAGGARLTAPIHAHFGVEDPRVSLDEAQAWADHTTG
                HFELRSHPGGHFYLNDQAPRVIEEIARVSSSVAS"
gene            4643. .7438
                /gene="olmRI"
CDS             4643. .7438
                /gene="olmRI"
                /codon_start=1
                /transl_table=11
                /product="luxR-family transcriptional regulatory protein"
                /protein_id="BAB69189.1"
                /db_xref="GI:15823972"
                /translation="MLIGRGQEIDRLRAAYSSVRSTGRSRAVLVESAAGCGKSEILED
                LVRHVESLGLIALRATAIAEERNTPLSVLRSLVNGAPFPDDLVRRFHTMSDAAARMIS
                TEGRDKARRDAAWYELMRTFCTEVRALAQRAPLVVAVDDIHHGDADSLRQLLHLARHC
                RNVPVLLTFTQSLAGGPADRTFGAELLRQPNLEHIRLPRFCRQDIADMLAEQGRADAG
                QGTVERYFAVSGGNPLLLRALLEDARLEPDDTVDAQDVGPRAGQAYGRAVLACLDRGG
                EKAHEVAAGLAVLGGAFTHELLSRLLDISAAALTENLRALRAAGIVVGTSFQHPAAKA
                AVLDSLEPSHRLRLHQRSAALLHASGAPAALAAEHLHEAGQADGAWAVGVLQQAAEQA
                LLADEARRAVSYLELAHAASADPRGRTEIKIRLGSVTRRVSAADAERHVDDALRGFGA
                GLLTPSDTGGLAALLLAHGRLEQARSVLEQGRRPGDGDADRTAARRQLAEIRDSVLNR
                GAWCAGTRAAWHENPAGRPSTGRAGTGLRAWLSGSHAALLAQRIAGTEGLLEITHLTD
                ATFDLVLNAVWTLLALGAVERALHWCDTFLREAKRRDAPGWEAVFAAVRAEGALRGGN
                LAEAERDARRCFAVLPERHRSALEGSAVSYLVTVQIARGKYEDAARQLSRPMADSLPD
                TIHWLDYLRARGLYYMATRQYHSALRDFHEVGRLAERWGTDWPALLPWRTDTAEALLR
                LGEREQGLRLATEQLAMVQDGAPRIKGISLRLQGLASEPAARLGLLNRAVVELSRADD
                RLELSRTLFDLADAHRETRQTRQAAMVLRRARQLAKSCGAQPPRDRIGSPDAAGASGE
                LFGEGPERNADWGHLSESERRVAGLAASGYTNREISVRLYITMSTVEQHLTRVYRKLK
                INGREELLAEARAGVQV"
gene            7658. .10486
                /gene="olmRII"
CDS             7658. .10486
                /gene="olmRII"
                /codon_start=1
                /transl_table=11
                /product="luxR-family transcriptional regulatory protein"
                /protein_id="BAB69190.1"
                /db_xref="GI:15823973"
                /translation="MLASAGQGEGGVVLVNAGPGVGRTALLDAFLRKATESGARVCAA
                AGSSAETEAEFGVVGQLFPPTGPLGAGFRLARAIGSRDAGPGLDADRLFDALCQELRR
                RPTVIAVDDAQLADAASLRFLRYLVRRLRTAPVMMVLTEPTGPDALPPSFQAELLRHP
                QCRRLGLLPLSRDGVTQVVSAQVGSDRVPRLAADFHAISGGNPLLVHALLADHRAAHR
```

```
Qy       601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
Db      8846 GTGGGTGGGGTGACGGTGATGTCGTCGCCGGTGACGTTGACGACGTTCAGTCGGCAGCGG 8905

Qy       661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
Db      8906 GGTTTGTCGGTGGATGGGCGGTGCAAGGCGTTCGCGGCTTCGGCGGATGGTTTTGGTGCT 8965

Qy       721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
Db      8966 GCCGAGGGTGTGGGTGTGTTGTTGGTGGAGCGGTTGTCGGATGCGCGGCGGTTGGGTCAT 9025

Qy       781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
Db      9026 CGGGTGTTGGCGGTGGTGCGGGGGAGTGCGGTCAATCAGGATGGTGCGTCCAATGGTCTG 9085

Qy       841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
Db      9086 GCGGCGCCGAATGGTCCGTCGCAGCAGCGGGTGATCCGTGCGGCGTTGGCTGACGCGGGT 9145

Qy       901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
Db      9146 CTGGCTCCTGCCGATGTGGATGTGGTGGAGGCGCATGGCACGGGGACGCGGTTGGGTGAT 9205

Qy       961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAG---CGGGCCGGTGACAGGCCG 1017
Db      9206 CCGATCGAGGCTCAGGCGTTGCTGGCGACGTATGGGCAGGGTCGTACCAGTGGGCGTCCG 9265

Qy      1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077
Db      9266 GTGTGGCTGGGGTCGGTGAAGTCGAACATCGGGCATACGCAGGCGGCGGCCGGTGTGGCT 9325

Qy      1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137
Db      9326 GGTGTGATGAAGATGGTGCTGGCGTTGGGTCGGGGTGTGGTGCCGAAGACGTTGCATGTG 9385

Qy      1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197
Db      9386 GATGAGCCGTCACCGCATGTGGACTGGTCGGCTGGTGAGGTGGAGTTGGCGGTTGAGGCG 9445

Qy      1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257
Db      9446 GTGCCGTGGTCGCG------GGGTGGGCGGGTGCGGCGGGCTGGTGTGTCGTCGTTCGGG 9499

Qy      1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
Db      9500 ATCAGCGGCACGAATGCGCATGTGATCGTGGAGGAGGCGCCTGCG 9544


RESULT 13
AR159871
LOCUS       AR159871               50937 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 1 from patent US 6251636.
ACCESSION   AR159871
VERSION     AR159871.1  GI:16222696
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 50937)
  AUTHORS   Betlach,M.C., Shah,S.Krishnakant., McDaniel,R. and Tang,L.
  TITLE     Recombinant oleandolide polyketide synthase
  JOURNAL   Patent: US 6251636-A 1 26-JUN-2001;
FEATURES             Location/Qualifiers
     source          1. .50937
                     /organism="unknown"
BASE COUNT     6672 a  16252 c  19273 g   8740 t
ORIGIN

  Query Match           52.9%;  Score 689;  DB 6;  Length 50937;
  Best Local Similarity   72.3%;  Pred. No. 2.9e-78;
  Matches  944;  Conservative    0;  Mismatches  340;  Indels   21;  Gaps    3;
```

```
Qy         1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
Db     13407 GAGCCGATCGCGATCGTGGGAATGGCATGCCGCTTCCCGGGAGGCATCGCTTCCCCGGAG 13466

Qy        61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
Db     13467 GATCTGTGGGACGTGCTCACCGCTGGTGAGGACGTTGTCTCCGGACTGCCGCAGAACCGC 13526

Qy       121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
Db     13527 GGGTGGGACTTGGGGCGCCTGTACGATCCCGATCCGGACCGGGCCGGTACGTCATACATG 13586

Qy       181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
Db     13587 CGTGAGGGTGCTTTCCTGCACGAGGCGGGGGAGTTCGACGCGGCCTTCTTCGGTATCTCG 13646

Qy       241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
Db     13647 CCGCGTGAGGCGTTGGCGATGGACCCGCAGCAGCGGTTGCTGCTGGAGACGTCCTGGGAG 13706

Qy       301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
Db     13707 GCCCTCGAACGGGCCGGCATCACTCCTTCCAAGCTGGCGGGCAGTCCGACCGGTGTGTTC 13766

Qy       361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
Db     13767 TTCGGCATGTCGAACCAGGACTACGCCGCCCAGGCGGGCGACGTGCCGTCCGAG------ 13820

Qy       421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
Db     13821 ------CTGGAGGGCTACCTGCTCACCGGCTCCATCTCCAGCGTCGCTTCGGGGCGTGTT 13874

Qy       481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
Db     13875 GCTTACACGTTCGGTCTTGAGGGGCCTGCGGTGACGGTGGATACGGCGTGTTCGTCGTCG 13934

Qy       541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
Db     13935 TTGGTGGCGTTGCATCTGGCGGTGCAGGGGTTGCGGCGGGGTGAGTGTTCGCTTGCGTTG 13994

Qy       601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
Db     13995 GTGGGTGGGGTGACGGTGATGTCGTCGCCGGTGACGTTGACGACGTTCAGTCGGCAGCGG 14054

Qy       661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
Db     14055 GGTTTGTCGGTGGATGGGCGGTGCAAGGCGTTCGCGGCTTCGGCGGATGGTTTTGGTGCT 14114

Qy       721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
Db     14115 GCCGAGGGTGTGGGTGTGTTGTTGGTGGAGCGGTTGTCGGATGCGCGGCGGTTGGGTCAT 14174

Qy       781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
Db     14175 CGGGTGTTGGCGGTGGTGCGGGGGAGTGCGGTCAATCAGGATGGTGCGTCCAATGGTCTG 14234

Qy       841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
Db     14235 GCGGCGCCGAATGGTCCGTCGCAGCAGCGGGTGATCCGTGCGGCGTTGGCTGACGCGGGT 14294

Qy       901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
Db     14295 CTGGCTCCTGCCGATGTGGATGTGGTGGAGGCGCATGGCACGGGGACGCGGTTGGGTGAT 14354

Qy       961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAG---CGGGCCGGTGACAGGCCG 1017
Db     14355 CCGATCGAGGCTCAGGCGTTGCTGGCGACGTATGGGCAGGGTCGTACCAGTGGGCGTCCG 14414

Qy      1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077
Db     14415 GTGTGGCTGGGGTCGGTGAAGTCGAACATCGGGCATACGCAGGCGGCGGCCGGTGTGGCT 14474

Qy      1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137
```

```
                     ASLVLAQALVDLGRVGEGPRLWLVTRGAVVAGPSDAGVVIDPVQAQVWGFGRVLGLEH
                     PELWGGLVDLPVGVDEEVCRRFVGVVASAGFEDQVAVRGSGVWVRRLVRAVVDGGGGG
                     WRPRGTVLVTGGLGGLGAHTARWLVGGGADHVVLVSRRGGSAPGAGDLVRELEGLGGA
                     RVSVRACDVADRVALRALLSDLGEPVTAVFHAAGVPQSTPLAEISVQEAADVMAAKVA
                     GAVNLGELVDPCGLEAFVLFSSNAGVWGSGGQAVYAAANAFLDALAVRRRGVGLPATS
                     VAWGMWAGEGMASVGGAARELSRRGVRAMDPERAVAVMADAVGRGEAFVAVADVDWER
                     FVTGFASARPRPLISDLPEVRAVVEGQVQGRGQGLGLVGEEESSGWLKRLSGLSRVRQ
                     EEELVELVRAQAAVVLGHGSAQDVPAERAFKELGFDSLTAVELRNGLAAATGIRLPAT
                     MAFDHPTATAIARFLQSELVGSDDPLTLMRSAIDQLETGLALLESDEEARSEITKRLN
                     ILLPRFGSGGSSRGREAGQDAGEHQDVEDATIDELFEVLDNELGNS"
misc_feature         650. .1906
                     /gene="oleAI"
                     /note="Region: ketoacyl synthase domain KS-Q"
misc_feature         2309. .3415
                     /gene="oleAI"
                     /note="Region: acyltransferase domain AT-L"
misc_feature         3485. .3724
                     /gene="oleAI"
                     /note="Region: acyl carrier protein domain ACP-L"
misc_feature         3806. .5056
                     /gene="oleAI"
                     /note="Region: ketoacyl synthase domain KS1"
misc_feature         5363. .6400
                     /gene="oleAI"
                     /note="Region: acyltransferase domain AT1"
misc_feature         7109. .7669
                     /gene="oleAI"
                     /note="Region: ketoacyl reductase domain KR1"
misc_feature         7943. .8200
                     /gene="oleAI"
                     /note="Region: acyl carrier protein domain ACP1"
misc_feature         8258. .9541
                     /gene="oleAI"
                     /note="Region: ketoacyl synthase domain KS2"
misc_feature         9848. .10882
                     /gene="oleAI"
                     /note="Region: acyltransferase domain AT2"
misc_feature         11723. .12274
                     /gene="oleAI"
                     /note="Region: ketoacyl reductase domain KR2"
misc_feature         12560. .12817
                     /gene="oleAI"
                     /note="Region: acyl carrier protein domain ACP2"
gene                 13118. .24568
                     /gene="oleAII"
CDS                  13118. .24568
                     /gene="oleAII"
                     /note="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="8,8a-deoxyoleandolide synthase 2"
                     /protein_id="AAF82409.1"
                     /db_xref="GI:9049536"
                     /translation="MTNDEKIVEYLKRATVDLRKARHRIWELEDEPIAITSMACHFPG
                     GIESPEQLWELLSAGGEVLSEFPDDRGWDLDEIYHPDPEHSGTSYVRHGGFLDHATQF
                     DTDFFGISPREALAMDPQQRLLLETSWQLFERAGVDPHTLKGSRTGVFVGAAHMGYAD
                     RVDTPPAEAEGYLLTGNASAVVSGRISYTFGLEGPAVTVDTACSSSLVALHLAVQALR
                     RGECSLAVVGGVAVMSDPKVFVEFSRQRGLARDGRSKAFAASADGFGFAEGVSLLLLE
                     RLSDARRLGHRVLAVVRGSAVNQDGASNGLAAPNGPSQQRVIRAALADAGLAPADVDV
                     VEAHGTGTRLGDPIEAQALLATYGQGRTSGRPVWLGSVKSNIGHTQAAAGVAGVMKMV
                     LALERGVVPKTLHVDEPSPHVDWSTGAVELLTEERPWEPEAERLRRAGISAFGVSGTN
                     AHVIVEEAPAEPEPEPEPGTRVVAAGDLVVPWVVSGRDAGALRAQAARLAAHVSSTGA
                     GVVDVGWSLVATRSVFEHRAVMVGTDLDSMAGSLAGFAAGGVVPGVVSGVAPAEGRRV
                     VFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDPLTGWSLVEVLRGGEAVLGRV
                     DVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRS
                     RAIARIAGGGGMVSVSLPAGRVRTMLDTYGGRLSVAAVNGPSSTVVSGDAQALDELLA
                     GCEREGVRARRVPVDYASHSAQMDQLRDELLEALADITPQHSSVPFFSTVTADWLDTT
                     ALDAGYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDTVEADAVA
                     LGSLRRDEGGLGRFLTSLAEAFVQGVPVDWSRTFEGASPRTVDLPTYPFQRQRFWLEG
                     SPALSSNGVEGEADVAFWDAVEREDSAVVAEELGIDAKALHMTLPALSSWRRRERQRR
                     KVQRWRYRVEWKRLPNSRAQESLQGGWLLVVPQGRAGDVRVTQSVAEVAAKGGEATVL
                     EVDALHPDRAAYAEALTRWPGVRGVVSFLAWEEQALAEHPVLSAGLAASLALAQALID
                     VGGSGESAPRLWLVTEAAVVIGAADTGAVIDPVHAQLWGFGRVLALEHPELWGGLIDL
                     PAVAGEPGSITDHAHADLLATVLATMVQAAARGEDQVAVRTTGTYVPRLVRSGGSAHS
                     GARRWQPRDTVLVTGGMGPLTAHIVRWLADNGADQVVLLGGQGADGEAEALRAEFDGH
                     TTKIELADVDTEDSDALRSLLDRTTGEHPLRAVIHAPTVVEFASVAESDLVRFARTIS
                     SKIAGVEQLDEVLSGIDTAHDVVFFSSVAGVWGSAGQSAYAAGNAFLDAVAQHRRLRG
                     LPGTSVAWTPWDDDRSLASLGDSYLDRRGLRALSIPGALASLQEVLDQDEVHAVVADV
                     DWERFYAGFSAVRRTSFFDDVHDAHRPALSTAATNDGQARDEDGGTELVRRLRPLTET
                     EQQRELVSLVQSEVAAVLGHSSTDAVQPQRAFREIGFDSLTAVQLRNRLTATTGMRLP
                     TTLVFDYPTTNGLAEYLRSELFGVSGAPADLSVVRNADEEDDPVVIVGMACRFPGGID
                     TPEAFWKLLEAGGDVISELPANRGWDMERLLNPDPEAKGTSATRYGGFLYDAGEFDAA
                     FFGISPREALAMDPQQRLLLETVWELIESAGVAPDSLHRSRTGTPIGSNGQFYAPLLW
                     NSGGDLEGYQGVGNAGSVMSGRVAYSLGLEGPAVTVDTACSSSLVALHLAVQALRRGE
                     CSLAIAGGVTVMSTPDSFVEFSRQQGLSEDGRCKAFASTADGFGLAEGVSALLVERLS
                     DARRLGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRAALADAGLAPADVDVVEA
                     HGTGTRLGDPIEAQALLATYGQGRAGGRPVVLGSVKSNIGHTQAAAGVAGVMKMVLAL
                     ERGVVPKTLHVDEPSPHVDWSAGEVELAVEAVPWSRGGRVRRAGVSSFGISGTNAHVI
                     VEEAPAEPEPEPGTRVVAAGDLVVPWVVSGRDAGALREQAARLAAHVSSTGAGVVDVG
                     WSLVATRSVFEHRAVMVGSELDSMAESLAGFAAGGVVPGVVSGVAPAEGRRVVFVFPG
                     QGSQWVGMAAGLLDACPVFAEAVAECAAVLDPVTGWSLVEVLRGGGEAVLGRVDVVQP
                     ALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAGGLSLADGARVVVLRSRAIAR
                     IAGGGGMVSVGLSAERVRTMLDTYGGRVSVAAVNGPSSTVVSGDVQALDELLAGCERE
                     GVRARRVPVDYASHSAQMDQLRDELLEALADITPQHSSVPFFSTVTADWLDTTALDAG
                     YWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVPGIEQTLDALDQNAAVLGSLR
                     RDEGGLDRLLTSLAEAFVQGVPVDWTHAFEGMTPRTVDLPTYPFQRQHYWPKPAPAPG
                     ANLGDVASVGLTAAGHPLLGAVVEMPDSDGLVLTGQISLRTHPWLADHEVLGSVLLPG
                     TAFVELAVQAADRAGYDVLDELTLEAPLVLPDRGGIQVRLALGPSEADGRRSLQLHSR
                     PEEAAGFHRWTRHASGFVVPGGTGAARPTEPAGVWPPAGAEPVALASDRYARLVERGY
                     TYGPSFQGLHTAWRHGDDVYAEVALPEGTPADGYALHPALLDAAVQAVGLGSFVEDPG
                     QVYLPFLWSDVTLHATGATSLRVRVSPAGPDTVALALADPAGAPVATVGALRLRTTSA

  Query Match           52.9%;  Score 689;  DB 1;  Length 24568;
  Best Local Similarity 72.3%;  Pred. No. 3.5e-78;
  Matches  944;  Conservative    0;  Mismatches  340;  Indels   21;  Gaps    3;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || |||||||| ||| | || |||||||||||| |||||||| || ||  |||||| ||
Db       8258 GAGCCGATCGCGATCGTGGGAATGGCATGCCGCTTCCCGGGAGGCATCGCTTCCCCGGAG 8317

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              || |||||||  ||  | || | || |  ||||   ||        | ||      |||||
Db       8318 GATCTGTGGGACGTGCTCACCGCTGGTGAGGACGTTGTCTCCGGACTGCCGCAGAACCGC 8377

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||  |  ||  |  |  |||  |  || ||    | | |||| ||||  ||  |
Db       8378 GGGTGGGACTTGGGGCGCCTGTACGATCCCGATCCGGACCGGGCCGGTACGTCATACATG 8437

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              |  |  || |  |||| |||| |||| | | |||||||| | |||||||| |||
Db       8438 CGTGAGGGTGCTTTCCTGCACGAGGCGGGGGAGTTCGACGCGGCCTTCTTCGGTATCTCG 8497

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              || |||||||  |||||||||| |||||||||||| |||||||||||||||||||||||
Db       8498 CCGCGTGAGGCGTTGGCGATGGACCCGCAGCAGCGGTTGCTGCTGGAGACGTCCTGGGAG 8557

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||  |||| ||||| || |||  |||       | |    ||    |  || || || |||
Db       8558 GCCCTCGAACGGGCCGGCATCACTCCTTCCAAGCTGGCGGGCAGTCCGACCGGTGTGTTC 8617

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
                 |||         |  |||||| | | |    || |  || | |||| |||
Db       8618 TTCGGCATGTCGAACCAGGACTACGCCGCCCAGGCGGGCGACGTGCCGTCCGAG------ 8671

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              |||||||  || | ||||||||  |||    |||||||   |||| ||| |
Db       8672 ------CTGGAGGGCTACCTGCTCACCGGCTCCATCTCCAGCGTCGCTTCGGGGCGTGTT 8725

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              || ||||  ||||| || || ||||| ||| | || ||||| |||| ||| ||| | |||
Db       8726 GCTTACACGTTCGGTCTTGAGGGGCCTGCGGTGACGGTGGATACGGCGTGTTCGTCGTCG 8785

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              | ||| || |||||||||||  ||| | |||||  |||||||| |||| ||| ||
Db       8786 TTGGTGGCGTTGCATCTGGCGGTGCAGGGGTTGCGGCGGGGTGAGTGTTCGCTTGCGTTG 8845
```

```
Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              || ||||||| ||||||| | |||||||||| | | | ||| ||||||||| |||||||
Db      70616 GGACTGTCGGCGGACGGCCGCTGCAAGGCGTTCGCCGATGCGGCCGACGGCGTCGGCTGG 70557

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              |||||||||||||| |||||  |  |||||||  ||||||| |||   ||    |||||
Db      70556 GGCGAGGGCGTCGGCATGCTCGTACTGGAGCGTCTGTCGGACGCGCGCCGCAACGGGCAC 70497

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              |||||||| |||||||| |||||||||||||| ||||||||||||||||| ||||| |||
Db      70496 CGGGTGCTCGCGGTGGTGCGCGGCAGTGCGGTGAACCAGGACGGTGCGTCCAATGGTCTG 70437

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              |||||||||||||| || || ||| |||||||||||||||||||| |||||| ||||||
Db      70436 ACGGCGCCGAACGGCCCCGCCCAGCAGCGGGTGATCCGGCAGGCGCTGGCGAGTGCGGGG 70377

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||||  | |||||||||||| | |||||||| |||||| ||||| |||||||||||| |||
Db      70376 CTGTCGGCGGCGGATGTGGACGCGGTGGAGGCGCACGGTACGGGTACGACGCTGGGCGAT 70317

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGG---GCAGCGGGCCGGTGACAGGCCG 1017
              |||||||||| ||||| |||||||||||||| ||    | ||||  | |  ||| |||||
Db      70316 CCGATCGAGGCCCAGGCGCTGCTCGCCACGTATGGCCAGGAGCGACCTGAGGACCGGCCG 70257

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077
              ||  |||||||||  | || |||||||||| ||   |   ||| ||  |||||||||
Db      70256 TTGCTGCTGGGGTCGGTCAAATCCAACATCGGTCATGCGCAGGCGGCTTCGGGTGTGGCG 70197

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137
              || |||||||||||||| ||||| |||| | || ||| |||| |||||| ||||||||
Db      70196 GGTGTCATCAAGATGGTGCTGGCGATGCGGCACGGTGTGCTGCCTCGGACGCTGCATGTG 70137

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197
              ||| | |||||| |||| || ||||||   || || || ||   |||||||||   | |
Db      70136 GATGAACCGTCGTCGCATGTCGACTGGAGTGCCGGTGCCGTCGAGCTGCTGACCTCCGAG 70077

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257
              |    |||||||  || || ||      || ||| || ||||| || || || ||||.|
Db      70076 GCCGAGTGGCCGCAGGGCGAGG------GGCCGCGCCGCGCGGGCGTCTCCTCCTTCGGC 70023

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGC 1295
              || |||| ||||| |||||||||||  ||||| || |
Db      70022 GTCAGCGGGACGAACGCGCATGTGATCCTGGAGCAGCC 69985
```

RESULT 12
AF220951

```
LOCUS       AF220951               24568 bp    DNA     linear   BCT 12-JUL-2000
DEFINITION  Streptomyces antibioticus 8,8a-deoxyoleandolide synthase 1 (oleAI)
            and 8,8a-deoxyoleandolide synthase 2 (oleAII) genes, complete cds.
ACCESSION   AF220951
VERSION     AF220951.1  GI:9049534
KEYWORDS    .
SOURCE      Streptomyces antibioticus.
  ORGANISM  Streptomyces antibioticus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 24568)
  AUTHORS   Shah,S., Xue,Q., Tang,L., Carney,J.R., Betlach,M. and McDaniel,R.
  TITLE     Cloning, characterization and heterologous expression of a
            polyketide synthase and P-450 oxidase involved in the biosynthesis
            of the antibiotic oleandomycin
  JOURNAL   J. Antibiot. 53 (5), 502-508 (2000)
  MEDLINE   20363406
   PUBMED   10908114
REFERENCE   2  (bases 1 to 24568)
  AUTHORS   Shah,S., Betlach,M. and McDaniel,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JAN-2000) Kosan Biosciences, Inc., 3832 Bay Center
            Place, Hayward, CA 94545, USA
FEATURES             Location/Qualifiers
.    source          1. .24568
                     /organism="Streptomyces antibioticus"
                     /db_xref="taxon:1890"
gene                 623. .13075
                     /gene="oleAI"
CDS                  623. .13075
                     /gene="oleAI"
                     /note="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="8,8a-deoxyoleandolide synthase 1"
                     /protein_id="AAF82408.1"
                     /db_xref="GI:9049535"
                     /translation="MHVPGEENGHSIAIVGIACRLPGSATPQEFWRLLADSADALDEP
                     PAGRFPTGSLSSPPAPRGGFLDSIDTFDADFFNISPREAGVLDPQQRLALELGWEALE
                     DAGIVPRHLRGTRTSVFMGAMWDDYAHLAHARGEAALTRHSLTGTHRGMIANRLSYAL
                     GLQGPSLTVDTGQSSSLAAVHMACESLARGESDLALVGGVNLVLDPAGTTGVERFGAL
                     SPDGRCYTFDSRANGYARGEGGVVVVLKPTHRALADGDTVYCEILGSALNNDGATEGL
                     TVPSARAQADVLRQAWERARVAPTDVQYVELHGTGTPAGDPVEAEGLGTALGTARPAE
                     APLLVGSVKTNIGHLEGAAGIAGLLKTVLSIKNRHLPASLNFTSPNPRIDLDALRLRV
                     HTAYGPWPSPDRPLVAGVSSFGMGGTNCHVVLSELRNAGGDGAGKGPYTGTEDRLGAT
                     EAEKRPDPATGNGPDPAQDTHRYPPLILSARSDAALRAQAERLRHHLEHSPGQRLRDT
                     AYSLATRRQVFERHAVVTGHDREDLLNGLRDLENGLPAPQVLLGRTPTPEPGGLAFLF
                     SGQGSQQPGMGKRLHQVFPGFRDALDEVCAELDTHLGRLLGPEAGPPLRDVMFAERGT
                     AHSALLSETHYTQAALFALETALFRLLVQWGLKPDHLAGHSVGEIAAAHAAGILDLSD
                     AAELVATRGALMRSLPGGGVMLSVQAPESEVAPLLLGREAHVGLAAVNGPDAVVVSGE
                     RGHVAAIEQILRDRGRKSRYLRVSHAFHSPLMEPVLEEFAEAVAGLTFRAPTTPLVSN
                     LTGAPVDDRTMATPAYWVRHVREAVRFGDGIRALGKLGTGSFLEVGPDGVLTAMARAC
                     VTAAPEPGHRGEQGADADAHTALLLPALRRGRDEARSLTEAVARLHLHGVPMDWTSVL
                     GGDVSRVPLPTYAFQRESHWLPSGEAHPRPADDTESGTGRTEASPPRPHDVLHLVRSH
                     AAAVLGHSRAERIDPDRAFRDLGFDSLTALELRDRLDTALGLRLPSSVLFDHPSPGAL
                     ARFLQGDDTRRPEPGKTNGTRATEPGPDPDDEPIAIVGMACRFPGGVTSPEDLWRLLA
                     AGEDAVSGFPTDRGWNVTDSATRRGGFLYDAGEFDAAFFGISPREALVMDPQQRLLLE
                     TSWEALERAGVSPGSLRGSDTAVYIGATAQDYGPRLHESDDDSGGYVLTGNTASVASG
                     RIAYSLGLEGPAVTVDTACSSSLVALHLAVQALRRGECSLALAGGATVMPSPGMFVEF
                     SRQGGLSEDGRCKAFAATADGTGWAEGVGVLLVERLSDARRLGHRVLAVVRGSAVNQD
                     GASNGLTAPNGPSQQRVIRAALADAGLVPADVDVVEAHGTGTRLGDPIEAQALLATYG
                     QGRAGGRPVVLGSVKSNIGHTQAAAGVAGVMKMVLALGRGVVPKTLHVDEPSAHVDWS
                     AGEVELAVEAVPWSRGGRVRRAGVSSFGISGTNAHVIVEEAPAEPEPEPERGPGSVVG
                     VVPWVVSGRDAGALREQAARLAAHVSGVSAVDVGWSLVATRSVFEHRAVMVGSELDAM
                     AESLAGFAAGGVVPGVVSGVAPAEGRRVVFVFPGQGSQWVGMAAGLLDACPVFAEAVA
                     ECAAVLDPLTGWSLVEVLRGGGEAVLGRVDVVQPALWAVMVSLARTWRYYGVEPAAVV
                     GHSQGEIAAACVAGGLSLADGARVVVLRSRAIARIAGGGGMVSVSLPAGRVRTMLEEF
                     DGRVSVAAVNGPSSTVVSGDVQALDELLAGCEREGVRARRVPVDYASHSAQMDQLRDD
                     LLEALATIVPTSANVPFFSTVTADWLDTTALDAGYWFTNLRETVRFQEAVEGLVAQGM
                     GAFVECSPHPVLVPGITETLDTFDADAVALSSLRRDEGGLDRFLTSLAEAFVQGVPVD
                     WSRAFEGASPRTVDLPTYPFQRQRYWLLDKAAQRERERLEDWRYHVEWRPVTTRPSAR
                     LSGVWAVAIPARLARDSLLVGAIDALERGGARAVPVVVDERDHDRQALVEALRNGLGD
                     DDLAGVLSLLALDEAPHGDHPDVPVGMAASLALVQAMADAAAEVPVWFATRGAVAALP
                     GESPERPRQALLWGLGRVVALEQPQIWGGLVDLPQHLDEDAGRRLVDVVGGLADEDQL
                     AVRASSVLARRLVRTPGHRMSSQAGGREWSPSGTVLVTGGTGALGAHVARWLAGKGAE
                     HLVLISRRGADAAGAAALRDSLTDMGVRVTLAACDAADRHALETLLDSLRTDPAQLTA
                     VIHAAGALDDGMTTVLTPEQMNNALRAKVTATVNLHELTRDLDLSAFVLFSSISATLG
                     IPGQANYAPGNSFLDAFAEWRRAQGLVATSIAWGPWSGGTGMAHEGSVGERLQRHGVL
                     AMEPAAAIAALDHTLASDETAVAVADIDWSRFFLAYTALRARPLIGEIPEARRMLESG
                     SGPGDLEPDRAEPELAVRLAGLTAVEQERLLVQLVREQAAVVLGHSGAEAVAPDRAFK
                     DLGFDSLTSVELRNRLNTATGLRLPVTAVFDYARPAALAGHLRSRLIDDDGDHGALPG
                     VEKHAIDEPIAIVGMACRFPGGIASPEDLWDVLTAGEDVVSGLPQNRGWDLGRLYDPD
                     PDRAGTSYMREGAFLHEAGEFDAAFFGISPREALAMDPQQRLLLETSWEALERAGITP
                     SKLAGSPTGVFFGMSNQDYAAQAGDVPSELEGYLLTGSISSVASGRVAYTFGLEGPAV
                     TVDTACSSSLVALHLAVQGLRRGECSLALVGGVTVMSSPVTLTTFSRQRGLSVDGRCK
                     AFAASADGFGAAEGVGVLLVERLSDARRLGHRVLAVVRGSAVNQDGASNGLAAPNGPS
                     QQRVIRAALADAGLAPADVDVVEAHGTGTRLGDPIEAQALLATYGQGRTSGRPVWLGS
                     VKSNIGHTQAAAGVAGVMKMVLALGRGVVPKTLHVDEPSPHVDWSAGEVELAVEAVPW
                     SRGGRVRRAGVSSFGISGTNAHVIVEEAPAEPSVEEGPGSVVGVVPWVVSGRDAGALR
                     AQAARLAAHVSSTGAGVVDVGWSLVATRSVFEHRAVMVGTDLDSMAGSLAGFAAGGVV
                     PGVVSGVAPAEGRRVVFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVLDRLTGWS
                     LVEVLRGGEAVLGRVDVVQPALWAVMVSLARTWRYYGVEPAAVVGHSQGEIAAACVAG
                     GLSLADGARVVVLRSRAIARIAGGGGMVSVGLSAERVRTMLDTYGGRVSVAAVNGPSS
                     TVVSGDAQALDELLAGCEREGVRARRVPVDYASHSAQMDQLRDELLEALADVTPQDSS
                     VPFFSTVTADWLDTTALDAGYWFTNLRETVRFQEAVEGLVAQGMGAFVECSPHPVLVP
                     GITETLDTFDADAVALSSLRRDEGGLDRFLTSLAEAFVQGVPVDWTHAFEGGRPRFVD
                     LPTYAFQRQRYWLHEEPLQEPVDEAWDAEFWSVVERGDATAVSDLLSTDAEALHTVLP
                     ALSSWRRRRVEHRRLQDWRYRVEWKPFPAALDEVLGGGWLFVVPRGLADDGVVARVVA
                     AVTARGGEVSVVELDPTRPDRRAYAEAVAGRGVSGVVSFLSWDDRRHSEHSVVPAGLA
```

REGTSASRLVFVTRNAVSADGGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADSG
DLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGGT
GGLGRELARHLVVERGVRHLLLVSRSGPEASGVGELRDELAAHGAEIAVRACDVADRA
AVEAVVGAVDAAHPLTAVIHTAGVLDDGVIASLTPERLSAVLRPKVDAAWNLHEATKD
LDLAAFVLYSSISGVLGSLGQGNYAAGNVFLDALAGHRAGLGLPAQSLAWGAWTQGGG
MTATLSDADMQRMASYGTAPLSPERGLALFDAATAGSATHLVAVGRVSGPVRAQGPVP
SMLRGLVKGARRTAVNTTGETAVSLADRLRAARPEERTQLMTDVVRTEAAAVLGHASS
QNVDARREFYELGFDSLTSVELRNRLSTVTGLRLSATVVFDSKNPTDLAARLYADLAA
QSALDGSASQAGFLPVAAPETDSLERMFLDALDSGKVPEAQRMLSALGALRPSFENTA
ELEDLPLPATLAEGPGTPRLICVSTPTANGGVHEYARLAASFRGERHVSALPLVGFAT
GERLPATAETAVRVIAESALRASDGNPFVLVGHSSAGAFAYLAAALLENTWGIRPEAV
VLLDTLSLRHEQNESIDYAGLMRRHFMVDEVSPVRMTNSRLSAMARWMGMLNQLEVRH
TTVPVLIIRATKETFGIGTDTGIYGEDHGSPVDVRSVDADHFSMVRDDAPETARIVRE
WLDSLGNA"

gene complement(9772. .15198)
/gene="pimS3"

CDS complement(9772. .15198)
/gene="pimS3"
/function="polyketide synthase"
/codon_start=1
/transl_table=11
/product="PimS3 protein"
/protein_id="CAC20920.1"
/db_xref="GI:12055071"
/translation="MTTSSETVVAALRASLKEAENLRRQNRKLVAAATEPIAIVAMAC
RYPGGVQSPEDLWELLTAGKDAIGAFPTDRGWDLDALRDAGVDRRGNAVSQEGGFLDG
VGDFDADFFGISPRESVTMDPQQRLLLETSWEAIERAGIDPTTLRGSRTGVFIGTNGQ
DYAYLLVRSLADATGDIGTGIAASATSGRLSYTLGLEGPAVTVDTACSSSLVALHTAS
HALRAGECSLALVGGVNVMSSPGSLMEFSRQGGLAADGRCKAFSDTADGTGWSEGVGV
LVVERLSDARRNGHPVLAVVRGSAVNQDGASNGFTAPSGPSQQRVIRQALAAAGLSAS
DVDVVEAHGTGTPLGDPIEAQALLATYGQDRDPERPLLLGSVKSNLGHTQAAAGAAGL
IKMILAMRHGVLPKTLYADVPSSHVDWTSGEVELLTEPQEWPKADRPWRAGVSSFGIS
GTNAHVIVEQAEPDEEPVAEPGLAPAVVPWLVSAKTEEALAAQLARISSRRGTASPVD
VGFSLASGRSAFEHRAVLLAQCPDGELSEAARGRAETSRPLAVLFSGQGAQRVGMGRE
LYVRFPVFAEALDEVLAHFDAGLRDVMFGDAEGLDETGFTQPALFAIEVALFRLAESL
GVRPDFVGGHSIGEIAAAHVAGVFSLTDACALVAARARLMQALPAGGAMVAVQASESE
VAARLVDGVSIAAVNGPEAVVLAGDETAVLRIAEELAAGGRKTQRLPVSHAFHSPLME
PMLEDFRRVAEGLSYGAPQIPLVSNVTGELAVEELVCSPDYWVRHVRETVRFADGVSA
LEAEGASAFLELGPDGVLTAMARHTLDETKVAVPVLRKDRPEETALLTALAHLHVAGI
GVDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVAGAGLRSAEHPLLGATLTL
AESGEVVLTGRLSLQTHPWLTDHAIGGNVFFPGTGFLELAMHAADQAGCDRVEDLVLM
TPLVLAEENGTQVQVLVGPADEAGTRTISIHARPEDDPDQPWMQHAGGTLTSGERVVD
FGSTVWPPQDAVAVDLEGFYEGTGYGPAFQGLRSVWSCGDGAYAEVALPDHVAEDAGA
FGLHPALLDAVLQSHRIAGVGSADDQFLPFAWRGVSLHAGGASVLRVRVVKTGDDSMS
IAAVDAEGAPVLSAEAMVMRAYRDPGTGQAAARHGAERDELLSLEWVSGPEVPPAGVR
CVSLGADELGVGASVASLADLSGDEDFVVLPVSGVGGDVPAAAHELTARTLGLLQEWS
GREGSSGSRLVFVTRNAVSADEGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADS
GDLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGG
TGGLGRELARHLVVERGVRHLLLVSRSGPEAPGVSGLSDELAAHGAEVTVRACDVADR
AAVHALVAAVPAEHPLTAVIHTAGVLDDGLVTSLTPERLSTVLRPKVDAAWNLHEATK
DLDLAAFVLYSSISGVLGAPGQANYAAGNVFLDALAWHRMGQGLPAQSLAWGAWTQDG
GMTATLSDTDMQRMASYGTAPLSRQRGLALFDAACARSAAHLVAVGRVGGPVRASAPV
PSVLRGLVKGARRTAVNTTGETAVSLADRLRAARPEERPRLMTDLVRTEAAAVLGHAS
AKNVDARREFNELGFDSLTSVELRNRLSALTGLRLSAALVFDYPNPTALADHFVSELV
EEAATGPALLAELERLDSALAAGAPDARTRAAVTARLAQILDDWRDAPAEEDGAEVAE
RLEAASTDEIFAFIDNELGRHSDR"

gene complement(15210. .43733)
/gene="pimS2"

CDS complement(15210. .43733)
/gene="pimS2"
/function="polyketide synthase"
/codon_start=1
/transl_table=11
/product="PimS2 protein"
/protein_id="CAC20921.1"
/db_xref="GI:12055072"
/translation="MDQEQKLRDYLKRASADLRRSRQRVGELEAAAREPIAIVGMSCR
FPGGVRSPEELWSLLAADGDGITKSPADRGWELLAEDAGEFSGGFLHDAAEFDASFFG
ISPREALAMDPQQRLLLESSWEAFERAGIDPAAVRGTRTGMFVGAMPQEYRVGPDDDV
QGFALTGTTTSVISGRLAYFFGAVGPAVTVDTACSSSLVALHLAAHSLRQGECSLALA
AGVTVMSSPTTFVEFNRQGGLSADGRCRSFADSADGTGWSEGVGVLVLERLSEARRNG
HEILAVIRGSAVNQDGASNGLTAPNGPSQRRVIEQALVSARLSADEVDVVEAHGTGTT
LGDPVEAQALLATYGQGRDDDRPLLLGSVKSNLGHTQAAAGMAGVIKMVLAMRHGVLP
RTLHVDAPSSHVDWTQGAVRLLTEHVPWPQGEQPRRAGVSSFGLSGTNAHTILEEAPA
APSPEAGQEDTAPVAEPLAAGAVPWLVSGRTREALRAQAARLLEHLTSRPGLAPLDVA
YSLATARSGLEHRAAFTAAGPDGARAALTALADGTPAPGLVQDTARTRSKLAFLFAGQ
GSQRPGMGRELAARFPVFATALDEVLGHFDEGLERPLKDVLFAAEGTPEAALLDQTGY
AQPALFAIEVALYRLAESFGITPDFLAGHSIGEIAAAHVAGVFTLADAAALVLARGRL
MQALPEGGAMVSLEATEEEVLPLLEDRPDRLSIAAVNGPAAVVIAGEADDVTAVAEHF
DALGRRTKRLRVSHAFHSPLMEPMLAEFRTAVSRLTPQAPVIPVVSGLTGTLATVEQL
TSPDYWLDHARHAVRFADALRLARRARRRHLPRMGPDGVLSAMAQSCCRGHGHRRRRR
LRHARRVRPGRPEAETLTAALAGLHTRGVAVQWEPYFQGTGARRADLPTYAFQRRRYW
PKSLPATGGDVRAAGLGAAHHPLLTAAVSVANSDGLLLTGRLSRRTHPWLADHAVRGT
VLLPGTAFLELAVRAGDEAGCGRVEELTLAAPLVLPEEGGVQVQVWVGSPDASGRRAV
SVHSRPDGPEELPWTQHAAGTLVAGEHHAGFDATVWPPADATPLDLDGFYARMADTGF
GYGPLFQGLRAAWRSGDDVYAEVALPDSDGAASTESSAAFGLHPALLDAALHAAAFVD
LGEDARGGLPFSWQDVTLHASGASTVRVKLTPDGDDAVALAVADTTGAPVASIGSLVL
RTLPDEQIDAAHSLVRDALFELRWSEARQTAHEAPASAAVLGPDLFGLERALAPTDLT
VTAVPGPDAAPHAPDVVLAPVAGTTEDDVVTSVHARTAHVLERLRTWLTDERYAASRL
VFVTRGALATDARQASDPVAAAVWGLVRAAQSEHPGRFGLLDLDPGTTEPATEPLLRA
LALSSDEPQTAVRGTAVLTARLARAQAPQAATEWDPEGTVVLTGATGGLGRVLARHLV
AERGVRHLLLVSRRGADADGAGELMAELAAHGAQVTLAACDLADRAAVRRLLSGVPAE
HPVTAVVHSAGSLDDGVIASLTPERISAVLRPKADAVWHLHEATRDLDLAAFVVFSSL
SGTVGAAGQGNYAAANAFLDALAQLRRAQGLPGLSLAWGPWAPTAGMTGGLTDDDLDR
LARMGTPALTEEQGIALFDAATATDAAVLLPTRIDLSVLRVQSEIPPLWRGLIRTPAR
RSVVSASETAVTLVQQLSRLAETDRREVVLDLVAGQVAAVLGHAGTADIDPRRPLREL
GFDSLTAVELRNRLSAATGLRTAATVIFDHPTVDALAAHLLDGLMGSEAAAAADAPRT

Query Match 54.5%; Score 710; DB 1; Length 84985;
Best Local Similarity 73.7%; Pred. No. 5.8e-81;
Matches 956; Conservative 0; Mismatches 315; Indels 27; Gaps 3;

```
Qy        1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
            |||||| |||  ||| ||||||||   |||||  | |||||||| ||| | || ||  ||
Db    71258 GACCCGGTCGTGATCGTCGGCATGAGCTGCCGCTACCCGGGCGGCGTCGGCTCGCCCGAG 71199

Qy       61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
            |||||||||  ||||  |  |||||  |||||  |     |||||||||||||||||||
Db    71198 GACCTGTGGCGCCTGGTGTCGGAGGGCGTGGACGCGGTGTCCGACTTCCCCACCGACCGT 71139

Qy      121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
            ||||||  |  |||    |  ||  ||  ||||||  |  ||  ||||||  |  ||  |
Db    71138 GGATGGGACGTGGAGAGCCTCTACAGCCCCGACCCCGAGGCGCTCGGCACCTCGTACACC 71079

Qy      181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
            |    ||  ||  |||||  |||||  ||||||  |  ||||||  |||  ||||||| ||  ||
Db    71078 CGCTCCGGTGGATTCCTCCACGAGGCGGCGGAGTTCGACCCCGATTTCTTCGGGATGAGC 71019

Qy      241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
            ||  ||  |||||  |||||||  ||  |  ||||||||||||||||||||  ||||||||
Db    71018 CCGCGCGAGGCGCTGGCGACCGACGCCCAGCAGCGGCTGCTGCTGGAGACGACCTGGGAG 70959

Qy      301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
            ||  |||||||  ||||  |||||  |||  ||||  |  ||  ||  ||||||||||||||||
Db    70958 GCCATCGAGCGCACGGGCATCGACCCGGCGTCGCTGCGGGGCAGCCGTACGGGCGTCTTC 70899

Qy      361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
            |||||||   |  |  ||||||||||   |  |  |  ||  ||
Db    70898 GCGGGCGTGATGTACACCGACTACGGCGACCTCCTCGTCGGCGACCAGT---------- 70850

Qy      421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
                |||||| ||  |  ||| |  | ||| |||| ||  || ||||  ||
Db    70849 -------TCGAGGGCTACCGCAGCAACGGCAGCGCGGCCAGCATCGCCTCCGGCCGGGTC 70797

Qy      481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              |||| ||||||  | || || ||||||  |||| ||||||||||  ||||||  | ||
Db    70796 TCGTACACCTTCGGTTTCGAGGGTCCGGCGGTCACGGTGGACACGGCATGCTCGTCGTCC 70737

Qy      541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
            || ||  | |||||   ||||  ||||||||||| ||||| ||||||  ||||||| |
Db    70736 CTGGTCGCCCTGCACTGGGCGGCGCAGTCGCTGCGCTCGGGCGAGTGCTCGCTCGCGGTC 70677

Qy      601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
            || ||||| ||  ||||  |||||||||  |    | |||| ||||||||| ||||| |||
Db    70676 GCGGGCGGTGTGACGGTGATGTCCACACCGACGACGTTCGTCGAGTTCTCGCGGCAACGC 70617
```

```
Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              ||||||||||||||| |||||  |  |||||||  |||||||  |||   ||    |||||
Db      10597 GGCGAGGGCGTCGGCATGCTCGTACTGGAGCGTCTGTCGGACGCGCGCCGCAACGGGCAC 10656

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              ||||||||  ||||||||  ||||||||||||||  ||||||||||||||||||  ||||| |||
Db      10657 CGGGTGCTCGCGGTGGTGCGCGGCAGTGCGGTGAACCAGGACGGTGCGTCCAATGGTCTG 10716

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||||||||||||  ||  ||  |||  ||||||||||||||||||||  ||||||  ||||||
Db      10717 ACGGCGCCGAACGGCCCCGCCCAGCAGCGGGTGATCCGGCAGGCGCTGGCGAGTGCGGGG 10776

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||||  |  ||||||||||||  |  ||||||||  ||||||  |||||  ||||||||||| |||
Db      10777 CTGTCGGCGGCGGATGTGGACGCGGTGGAGGCGCACGGTACGGGTACGACGCTGGGCGAT 10836

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGG---GCAGCGGGCCGGTGACAGGCCG 1017
              ||||||||||| ||||||  ||||||||||||||  ||    |  ||||   |  |  |||  |||||
Db      10837 CCGATCGAGGCCCAGGCGCTGCTCGCCACGTATGGCCAGGAGCGACCTGAGGACCGGCCG 10896

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077
               ||  |||||||||  |  ||  ||||||||||||  ||  |   |||  ||  |||||||||
Db      10897 TTGCTGCTGGGGTCGGTCAAATCCAACATCGGTCATGCGCAGGCGGCTTCGGGTGTGGCG 10956

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137
              ||  ||||||||||||||  |||||  |||||  |  ||  |||  ||||  ||||||  ||||||||
Db      10957 GGTGTCATCAAGATGGTGCTGGCGATGCGGCACGGTGTGCTGCCTCGGACGCTGCATGTG 11016

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197
              |||  |  ||||||  ||||  ||  ||||||    ||  ||  ||  ||    |||||||||      |  |
Db      11017 GATGAACCGTCGTCGCATGTCGACTGGAGTGCCGGTGCCGTCGAGCTGCTGACCTCCGAG 11076

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257
              |    |||||||  ||  ||  ||        ||  |||  ||  |||||  ||  ||  ||  |||||
Db      11077 GCCGAGTGGCCGCAGGGCGAGG------GGCCGCGCCGCGCGGGCGTCTCCTCCTTCGGC 11130

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGC 1295
              ||  ||||  |||||  ||||||||||  |||||  ||  |
Db      11131 GTCAGCGGGACGAACGCGCATGTGATCCTGGAGCAGCC 11168
```

RESULT 11
SNA278573/c

```
LOCUS       SNA278573              84985 bp    DNA     linear   BCT 17-MAY-2001
DEFINITION  Streptomyces natalensis pimaricin biosynthetic gene cluster.
ACCESSION   AJ278573
VERSION     AJ278573.1  GI:12055067
KEYWORDS    ABC transporter; cholesterol oxidase; cytochrome P450 monoxygenase;
            ferredoxin; glucosyl transferase; macrolide efflux pump; mycosamine
            dehyralase; orfX; pimA gene; pimB gene; pimC gene; pimD gene; pimE
            gene; pimF gene; pimG gene; pimH gene; pimI gene; pimJ gene; pimK
            gene; pimS0 gene; pimS1 gene; pimS2 gene; pimS3 gene; pimS4 gene;
            polyketide synthase; sensory transduction protein; thioesterase.
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 84985)
  AUTHORS   Aparicio,J.F., Fouces,R., Mendes,M.V., Olivera,N. and Martin,J.F.
  TITLE     A complex multienzyme system encoded by five polyketide synthase
            genes is involved in the biosynthesis of the 26-membered polyene
            macrolide pimaricin in Streptomyces natalensis
  JOURNAL   Chem. Biol. 7 (11), 895-905 (2000)
  MEDLINE   20547809
   PUBMED   11094342
REFERENCE   2  (bases 1 to 84985)
  AUTHORS   Aparicio,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2000) Aparicio J.F., Molecular Genetics,
            Institute Of Biotechnology - Inbiotec, AV. REAL 1 - LEON, 24006,
            SPAIN
FEATURES             Location/Qualifiers
     source          1. .84985
                     /organism="Streptomyces natalensis"
                     /db_xref="taxon:68242"
CDS                  <1. .2086
                     /note="ORFX"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAC20917.1"
                     /db_xref="GI:12055068"
                     /translation="TRPDAFREATEPYGALLAPLMPERSADRVEGTKWTSDWSQARFL
                     AHDAVCEVLLTLAGQHPLVLLLEDLQWADTASLDLLRLLSPRCQGHPLGIVLTARKFE
                     TESDATLRRIVSDVLRGPRTATLRLCGLPRQAVGALVEAHAGSGVDPEVVEVLHRRSE
                     GNPYFVMQLLSLVGDARRLHRPDAVDALLAHIPTGVREALNQRFAALPEPVLRVLRLC
                     AVIGAEVDTDLLRRTATCDASVNAGLESAIRAGLLGEDPHHPGQLHFAHALVQETLFD
                     ELPRADRRRLHAKVADALYARRRGHVGDEEIERVAHHSWHAKSALPAAEVLPRLLLAA
                     EHAEQAMAYEQVEIWLRRAVHLVGFLPPDDPATVRLEQKLHIQLGQVLATTRGYGHPE
                     AETALTRGRALSTVTHAPEDPSVLWALCAASLVTGRYDASRQLSGLLRDLSWQTREPV
                     AVLGAAYGEGIVLHVRGMLPQALAELEHGVERADHFANEGHALARTFQHDPRVSCRSY
                     NTFTHWLLGHRQIADARRRELLSLTEYESRPSDRSFALYVDGVVAAWEGDTRTALASS
                     DEGVRVAGDHGLLYWKAMLGVVKGWGLVHAGDRDEGLALMHTSLAELGQSRTYLRHPL
                     HLGLLGQAQHHNGQIEEARTTFRMLLAAVEQRRERVYLHPALPATPLLHELLGHGADE
                     ALADRDPGRGQV"
gene                 2190. .3566
                     /gene="pimK"
CDS                  2190. .3566
                     /gene="pimK"
                     /function="putative glucosyl transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="PimK protein"
                     /protein_id="CAC20918.1"
                     /db_xref="GI:12055069"
                     /translation="MESARRPILFVSLPESGLLNPLLVLAGELSRQGVEDLWFATDEP
                     RRNDVKRIAEGSPVEFASLGEVDSEMSAVTWSDEVYREVTQPSRFKAHRAVVRHTYRP
                     GLQAEKFRRLQAVIDEVQPALMVIDCISGFAVDAAIARNIPYVLSVPFLPSNVLTAHT
                     HFAKSYTPRGFPVPHTGLSRRMTLAQRVANELFKLRTFAMFLNPRLGKVLAEDNRRRN
                     ELGLPKASFMARIEHADLVLCNSLAELDYPFDIPEKMRLVGAMVPPLPEAPDDQDLSR
                     WLDAQSSVVYVGLGTITRLTREQVGSMVEVARRLEDRHQVLWKLPSEQQHLLPPRESL
                     PGNLRVESWVPSQMDVLAHPHVKVFFTHGGGNGFNEGMYFGKPLVVRPLWVDCYDQAV
                     RGQDFGLSLTLDRPQTIDVNDVVDKLTRVLGTPSFYEKAERRAALMRSAGGRETAAGL
                     VLSLPALA"
gene                 complement(3662. .9736)
                     /gene="pimS4"
CDS                  complement(3662. .9736)
                     /gene="pimS4"
                     /function="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="PimS4 protein"
                     /protein_id="CAC20919.1"
                     /db_xref="GI:12055070"
                     /translation="MPNDKPNDKKLVDYLKWVTNDLHQTRQRLREVESAKQEPIAIVG
                     MSCRLPGGVRSPEEFWQLLKDGRDGISTFPTDRGWDLDLLAGDGSGSSATAEGGFIDA
                     ASFDAGFFGISPREAVTMDPQQRILLESSWEALERAGIDPVSLRGTETGVFVGTSGID
                     YINTVMNSREDIEGHGSTGLIASILSGRVSYTFGLEGPAVTVDTACSSSLVSLHLAAQ
                     ALRVGECSLALAGGVTAISTPMSFVGFSRQGGLAPDGRCKAFSDTADGTAWSEGVGVL
                     VVERLSDARRNGHPVLAVVRGSAVNQDGASNGLTAPSGPSQQRVIRQALAAAGLSASD
                     VDVVEAHGTGTALGDPIEAQALLATYGRDRDPERPLLLGSVKSNLGHTQAAAGAAGLI
                     KMILAMRHGVLPRTLHVEAPSSHVDWSSGALQLLTEPQEWPKADRPWRAGVSSFAISG
                     TNAHIILEQPPAAEEPEDEEGPHTSPSAVPWLVSAKSEEALTAQLDRISALTDRSPLD
                     LGLSLATTRAHLEHRAVLLSEGGETVEVARGHAEHTAGKLAVLFSGQGAQRVGMGREL
                     YGRFPVFAEALDGVLAHFDAGLRDVMFGEAEGLDETGFTQPALFAIEVALFRLSESLG
                     VRPDFVAGHSIGEIAAAHVAGVFSLADACALVAARARLMQALPAGGAMVAVQAAESEV
                     AARLVDGVSIAAVNGPEAVVIAGEEREVLRIAEEFTGLGRKTRRLPVSHAFHSPLMEP
                     MLEDFRRVAEGLSYEAPRIALVSNLTGELATEDLVRSAEYWVRHVRETVRFADGVSAL
                     AAEGASVFLELGPDGVLTAMAQHTLDGTATTVPALRKDRSEETALLTALAQLHVAGVG
                     VDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVVGAGLRSAEHPLLGATLTLA
                     ESGEVVLTGRLSLQAYPWLTDHTMGGSVLFPATGFLELAMRAADQVGCDRVDEFVLMT
                     PLVLAEESGTQVQVLVGPADEAGTRTISIHARPDGDPDQPWMQHAGGTLTSGERIVDF
                     GSTVWPPQGAEAVDLEGFYDTTGYGPAFQGLRSVWQCEDGAYVEVALPDQVAEDAGAF
                     GLHPVLLDAVLQSRRMAGIGAAGDTFLPFAWRGVSLHAGGASVLRARVIKTGEDTLSL
                     VAVDAEGAPVLSVEELVLRTQLTHTDAPLTARRGAERDALLSLEWVSGPEVPPAGVRC
                     VSLGADELGVGASVATLADLSGDEDFVVVPVSGVGGDVPAAAHELTARTLGLLQEWSG
```

```
VQASTAALDEPVAIIGMACRFPGGVHSPEALWRLLAEGGDAITPMPADRGWDLDRLYH
PDPDHQGTSYARGGGFLDGAADFDADFFGISPREALAMDPQQRLLLETWEVLEQAGID
PESLRGSSTGVFAGTNTQDYGTALDAAQDEAGGHRLTGNAMSVVSGRVSYTFGFEGPA
LTVDTACSSSLVALHMAAQALRQGECSLAVAGGVTVMATPSSFVEFARQRGLAPDGRC
KPFAAAADGTGWSEGVGLLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLSAPSGP
SQQRVIRQALANARVAASEVDAVEAHGTGTTLGDPIEAQALLATYGQERPLLLGAVKS
NLGHTQAAAGVAGVMKMVLAMRHGMLPRTLHVDEPTGHVDWTAGAVELLTEHTDWPET
GHPRRAAVSAFGISGTNAHVVLELPAAEQPLVEQPSAAEPDAPATAPDRTPTASDGTA
PLLLSAKSESALRAQAARLHSHLERDPALRLTDAAYTLMTHRTAFAHRAAVRAADHEA
ALRALTALAAGEADPAVDTGTAHTGRDAVLFSGQGSQRIGMGRELSGRYPVFAEAFDT
VCAALDEHLDRPLRDVVRGEDEELLNRTVYAQAGLFAIEVALFRLVESWGVRPHYVAG
HSVGEIAAAHVAGVFSLADACALVAARGRLMQALPAGGAMAAIRATEDEVLPHLADSV
SIAAVNGPSSVVVSGAEHAVLSIAAHFEGAGRKTTRLRVSHAFHSPLMDPMLADFRAV
AEGLTYGEPELAVVSNVTGQLATPDQLRTPEYWVTHVRAAVRFADGIRALGAEGVTRF
LELGPDGVLSALARESAPDDAVCTPVLRKDRSEAATLLAALTHLHVHGTEIDWTAFLA
GRDAHAVDLPTYAFQHQRFWPTPDHTRTGDLGAVGLEATGHPLLSAAVELPDGEGLLF
TTRLSLQTHPWLAGHVVMGSVLLPGTAFAELALRAADEVGCDRVDELTLAAPLVLPEH
GGVQLQLRVGPADASGRRTLTARSRAEGDGDRPWVQHATGVLAEGESTPEPGYDFHTE
SWPPADAAPVELSGLYPDFAAHGFDYGPHFQGLRTAWRRGDEVFAEVALPAEAEGEAS
AYGLHPALLDAALHVVAFNGVDRGVVPFSWESVALHATGASAVRIRVVRHSGDTVSVD
VADTTGEPVASIGTLVLRAVSADQLAGGADPAVRDALFRVQWNPVRLPPAGAAVTVAT
LGSLAGAPFDGYPDLASLARSGRVAGAVLVPVEAGAGEVVADDVVGATHATAARALDL
ARSWLADDRFAASRLVFVTRGAVSGADLAGAAVWGLVRSALSEHPGRFGLVDLDDDAE
LALVPRVLASDEPQLLVRGGEVLAARLARAQSSHAVTWDPSGTVLVTGGTGGLGRVMA
RHLVVEHGVRNLLLVSRRGPAAEGAEELVTELRHSGAEVAVEACDVTDAAAVADLVAR
HRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRGLDLDAFVVFSSV
AGTIGSPGQANYAAGNAFLDALAHHRRAAGLPAASLAWGPWSRDGGMTGTLTDVDSSA
SPGRHARTHPRTGRGLFDAALAAGDAHLLPVRFDWASLRAQGEVPPLLRGLIRTRARR
SAVGGSAAAAGLVGRLSGRGTVERREVLLDLVRAQIAVVLGHANPETIESTRVFQDLG
FDSLTAVELRNRLNNATGLRLSATAVFDYPTADALVDFLLDELFGAQEEAELPAPVPS
PAGAADDPVVIVGMSCRYPGGVGSPEDLWRLVSEGVDAVSDFPTDRGWDVESLYSPDP
EALGTSYTRSGGFLHEAAEFDPDFFGMSPREALATDAQQRLLLETTWEAIERTGIDPA
SLRGSRTGVFAGVMYTDYGDLLVGDQFEGYRSNGSAASIASGRVSYTFGFEGPAVTVD
TACSSSLVALHWAAQSLRSGECSLAVAGGVTVMSTPTTFVEFSRQRGLSADGRCKAFA
DAADGVGWGEGVGMLVLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPAQQR
VIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLATYGQERPEDRPLLLGSVKS
NIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHVDWSAGAVELLTSEAEWPQG
EGPRRAGVSSFGVSGTNAHVILEQPGPDAADAAPDATVTDPGALAWVLSARNEAALRC
QAARLLSLVAGSDALCARDIGHSLVTGRSSFAHRAVVWGQDRDALVRALSALAVGEAD
AGLAEGASGAGRTAFLFSGQGSQRLGMGWELYARYPVFADAFDAVCAALDEHLERPLR
DVVWGEDAELLNQTAYAQAGLFAIEVALYRLAESWGMRPDFVAGHSIGEVAAAHVSGV
FSLPDACALVAARGRLMQQLPSGGAMMAIRATEDEVLPHLAEGVSLAAVNGPSSVVIS
GAEDAVLAIAAHFAGEGRKTTRLRVSHAFHSPLMEPMLEEFRAVVTRLSFGTPTIPVV
SNLTGRLAEPEQLAHADYWVRHVREAVRFADGIQALRAEGVTRFLELGPDGVLSAMAR
ESASDDAVLAPVLRRDRPEETALLGALAQLYVRGAHVDWTVPFAGSGARWADLPTYAF
QHERFWPSGGVARPGDVRSAGLGSAGHPLLGAAVELAGSGGLLFTGRLSVSSHPWLAD
HVVLGSVLVPGTALVELVLRAADEAGCDLLEELTLAAPLVLPASGAAVQVQVAVGEPD
EAGRRPVSVHAREGEGPWTLHASGAVTSGAEVPPFDATVWPPKGAEPVDVADCYDVLA
DAGLTYGPAFHGLQAAWKLGGDVYAEAKLPESTDGDAYGLHPALFDAALHASALGGAE
AGGVPFSWAGVSLHATGASHLRVRIREAGGALSVAIADTSGAPVASVESLVIRPLSAG
QVQAADRDALFKADWVPVPLTDERVEPGTGPEGEPLRTYADLDSLEGAAVPGTVLVAP
PSGAAGTVESVHAATVWALEMVQAWLADDRFATSRLVFVTRGAAFGADLAAAAVRGLV
RSAQSENPGRFGLVDMDGDADTTVPAQALATDEPELLVRGGEVLAARLVRAQSSHTVT
WDPSGTVLITGGTGGLGRSVARHLVSEHGVRSLLLVSRRGPAAEGAGELVAELRGSGA
EVVIEACDVTDAVAVADLVARHRISAVVHTAGVLDDGVVESLTPERLAVVLRPKVDAA
WNLHEATRGLDLDAFVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLA
WGPWSQDGGMTGTLSDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLA
ALRAQGEPQPLLRGLIRTRTRRSGAAAASGIAQRLAGLSTAERREALLDVVRAQIATV
LGHAGPETIAPDRAFQDLGLDSLTAIELRNLLGKATGLRLPATTVFDYPTVDALAAHL
LDELFGAETGTATETPLPVPGLPSLADDPVVIVGMSCRFPGGVASPEDLWRLVADGVD
AVSAFPTDRGWEIDDTYDPEREGAIATRSGGFLHDAAEFDPEFFGMSPREALTTDAQQ
RLLLETTWEALERAGMDPATLRGSRTGVFAGVMYHDYSTLLSGREFEGYQGSGSAGSV
ASGRVSYTFGFEGPAVTVDTACSSSLVALHLAAQSLRSGECSLALAGGVTVMSTPLTF
VEFSRQGGLSADGRCKAFADAADGVGWAEGAGILVLERLSDARRNGHRILATVRGSAV
NQDGASNGLTAPNGPAQQRVIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLA
TYGQERPEDRPLLLGSVKSNIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHV
DWSAGAVELLTSEAEWPQGEGPRRAGVSSFGISGTNAHVILEQPEPVAAETESITPDT
APDAAEDEAADSGTPVPALLSGRSASALRAQAARLLSRLDGDPGPRITDVAYSLATGR
SAFPHRAVILAANRADLLHSLSALAEGHTEAPAVVAQDRARSGKLAFLFSGQGSQRLG
MGRELYGRYPAFAEALDAVCAALDAHLDRPLRDVIWGEDAELLNRTGYAQTGLFAIEV
ALFRLLESWGVRPDHLLGHSIGEIAAAHVAGVLSLPDACALVAARGRLMQQLPSGGAM
MAIRATEDEVLPHLAEGVSLAAVNGPSSVVVSGAEDEVLALAAHFEEEGRKTTRLRVS
HAFHSPLMEPMLADFRAVADGMTYAAPRIPVVSNVTGRPATAEELCCAEYWVGHVREA
                    VRFADGVGALREQGVTTFLELGPDGSLSALAAESAADDSVLAPVLRKNRPEAPALLTA
                    LARLHAQGTPVDWSAAFAGTGARWVDLPTYAFQHERFWPSGGAARAGDVRSAGLGSAG
                    HPLLGAAVELAGSGGRLLTGRLSLSSHPWLADHVVLGSVLVPGTALMELVLRAADEVD
                    CAAVDELTLAAPLVLPASGAAIQVQVWVGEPDEAGRRPVSVHAREGEGPWTLHADGAL
                    APAAETVPFDTAIWPPQGAEHLDAAGCYERFADAGFAYGPVFQGLRAAWKLGEDIYAE
                    VALPEGTDGNAYGLHPALFDAALHAALLGGEGTDEAAVPFSWNGVTLHATGASRVRVR
                    IRPTEGGTSIALVDTAGAPVASVRSLTARPITAGQLQTGDRDSLFQVDWTTLHLTDER
                    ANSLALLGKDTEGILDTLSLQPHADLDDLAATGVHDTVLAPLPTRTAGTVESVHAATT
                    GALALIRSWLADDRFAASRLVFVTRGAVSGTDLAGASVWGLVRSALLEHPGRFGLVDV
                    DVDQDAEVPLVPRALASDEPQVLVRGGEVLAARLVRAQSSDTVTWDPSGTVLITGGTG
                    GLGRSVARHLVSEHGVRSLLLVSRRGPAAEGVDALVAELAECGAQVTVEACDVTDAVA
                    VADLVARHRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRGLDLDA
                    FVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLAWGPWSQDGGMTGTL
                    SDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLAALRAQGEPQPLLRG
                    LIRTPGRRTAAAATEGDTAAAFAGRLTGLSAAEGREVVLGAVRSQIAGVLGHAEATEI
                    DQDRAFLDLGFDSLTAVELRNRLGAVTGIRLPATLLFDYPTPAELVAHLHARIAPEPT
                    VGPEALLGELERMEKSFGGLDLTEEMHEQIAGRLEVLRAKWDALRDTAAAAGHDGSPS
                    DEDFDFESASDDEVFDLLDNELGLS"
BASE COUNT     2611 a    7486 c    7397 g    2900 t
ORIGIN

  Query Match             54.5%;  Score 710;  DB 6;  Length 20394;
  Best Local Similarity   73.7%;  Pred. No. 8.5e-81;
  Matches  956;  Conservative    0;  Mismatches  315;  Indels   27;  Gaps    3;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              |||||| |||  ||| ||||||||   |||||.| |||||||| ||| | || ||  ||
Db       9895 GACCCGGTCGTGATCGTCGGCATGAGCTGCCGCTACCCGGGCGGCGTCGGCTCGCCCGAG 9954

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              |||||||||   ||||   |  |||||  |||||  |       ||||||||||||||
Db       9955 GACCTGTGGCGCCTGGTGTCGGAGGGCGTGGACGCGGTGTCCGACTTCCCCACCGACCGT 10014

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              ||||||  |  |||  |   ||  ||  ||||||  |  || |||||| |  ||  |
Db      10015 GGATGGGACGTGGAGAGCCTCTACAGCCCCGACCCCGAGGCGCTCGGCACCTCGTACACC 10074

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              |    || || ||||| ||||| |||||| | |||||| |||  |||||||| || ||
Db      10075 CGCTCCGGTGGATTCCTCCACGAGGCGGCGGAGTTCGACCCCGATTTCTTCGGGATGAGC 10134

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              || || ||||| |||||||  ||  | ||||||||||||||||||||||| ||||||||
Db      10135 CCGCGCGAGGCGCTGGCGACCGACGCCCAGCAGCGGCTGCTGCTGGAGACGACCTGGGAG 10194

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              || |||||||  |||| ||||| ||| |||| | || ||  |||||||||||||||||
Db      10195 GCCATCGAGCGCACGGGCATCGACCCGGCGTCGCTGCGGGGCAGCCGTACGGGCGTCTTC 10254

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              |||||||   | | |  ||||||||||   |   | | |  ||    ||
Db      10255 GCGGGCGTGATGTACACCGACTACGGCGACCTCCTCGTCGGCGACCAGT---------- 10303

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
                | ||||||  ||  |  ||  ||| | | |||  |||| ||   ||  ||||  ||
Db      10304 -------TCGAGGGCTACCGCAGCAACGGCAGCGCGGCCAGCATCGCCTCCGGCCGGGTC 10356

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              | ||||  |||||  | ||  ||  |||||  |||| |||||||||  |||||  | ||
Db      10357 TCGTACACCTTCGGTTTCGAGGGTCCGGCGGTCACGGTGGACACGGCATGCTCGTCGTCC 10416

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              || ||  | |||||   ||||   |||||||||||| ||||| ||||||  |||||||  |
Db      10417 CTGGTCGCCCTGCACTGGGCGGCGCAGTCGCTGCGCTCGGGCGAGTGCTCGCTCGCGGTC 10476

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              || ||||| ||  |||| ||||||||  |      | |||| ||||||||||  ||||| |||
Db      10477 GCGGGCGGTGTGACGGTGATGTCCACACCGACGACGTTCGTCGAGTTCTCGCGGCAACGC 10536

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              || ||||||| |||||| | ||||||||| | | ||| ||||||||  |||||||
Db      10537 GGACTGTCGGCGGACGGCCGCTGCAAGGCGTTCGCCGATGCGGCCGACGGCGTCGGCTGG 10596
```

```
  Best Local Similarity   73.7%;  Pred. No. 8.5e-81;
  Matches  956;  Conservative    0;  Mismatches  315;  Indels   27;  Gaps    3;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              |||||| ||| ||| |||||||| ||||| | ||||||||| ||| | || || ||
Db       9895 GACCCGGTCGTGATCGTCGGCATGAGCTGCCGCTACCCGGGCGGCGTCGGCTCGCCCGAG 9954

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              ||||||||| |||| |  |||||  ||||| |        ||||||||||||||
Db       9955 GACCTGTGGCGCCTGGTGTCGGAGGGCGTGGACGCGGTGTCCGACTTCCCCACCGACCGT 10014

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              ||||||  |  |||    |   ||  ||  ||||||  |  || ||||||  |  ||  |
Db      10015 GGATGGGACGTGGAGAGCCTCTACAGCCCCGACCCCGAGGCGCTCGGCACCTCGTACACC 10074

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              |     || || |||||  |||||  ||||||  |  ||||||  |||  ||||||||  || ||
Db      10075 CGCTCCGGTGGATTCCTCCACGAGGCGGCGGAGTTCGACCCCGATTTCTTCGGGATGAGC 10134

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              || || |||||  |||||||  ||  | ||||||||||||||||||||||||  ||||||||
Db      10135 CCGCGCGAGGCGCTGGCGACCGACGCCCAGCAGCGGCTGCTGCTGGAGACGACCTGGGAG 10194

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||  |||||||  |||| |||||  |||  |||| |  || ||   ||||||||||||||
Db      10195 GCCATCGAGCGCACGGGCATCGACCCGGCGTCGCTGCGGGGCAGCCGTACGGGCGTCTTC 10254

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              |||||||   | | |  ||||||||||    |  | |  ||     ||
Db      10255 GCGGGCGTGATGTACACCGACTACGGCGACCTCCTCGTCGGCGACCAGT---------- 10303

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
                | |||||| ||  |  || ||| | | ||| |||| ||   || |||||  ||
Db      10304 -------TCGAGGGCTACCGCAGCAACGGCAGCGCGGCCAGCATCGCCTCCGGCCGGGTC 10356

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              | |||| ||||||  | || || |||||| |||| ||||||||||  ||||||  | ||
Db      10357 TCGTACACCTTCGGTTTCGAGGGTCCGGCGGTCACGGTGGACACGGCATGCTCGTCGTCC 10416

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              || ||  | |||||  ||||  |||||||||| ||||| |||||| |||||||||  |
Db      10417 CTGGTCGCCCTGCACTGGGCGGCGCAGTCGCTGCGCTCGGGCGAGTGCTCGCTCGCGGTC 10476

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              || ||||| ||  |||| |||||||| |      | |||| |||||||||| ||||| |||
Db      10477 GCGGGCGGTGTGACGGTGATGTCCACACCGACGACGTTCGTCGAGTTCTCGCGGCAACGC 10536

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              || ||||||| ||||||| | |||||||||| | | | ||| |||||||||  |||||||
Db      10537 GGACTGTCGGCGGACGGCCGCTGCAAGGCGTTCGCCGATGCGGCCGACGGCGTCGGCTGG 10596

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCCGGATCGCGGTCGGCTGGGGCAT 780
              |||||||||||||| |||| |  |||||||  |||||||  |||    ||    |||||
Db      10597 GGCGAGGGCGTCGGCATGCTCGTACTGGAGCGTCTGTCGGACGCGCGCCGCAACGGGCAC 10656

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              |||||||| |||||||| |||||||||||||| |||||||||||||||||| ||||| |||
Db      10657 CGGGTGCTCGCGGTGGTGCGCGGCAGTGCGGTGAACCAGGACGGTGCGTCCAATGGTCTG 10716

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              |||||||||||||| || || ||| |||||||||||||||||| ||||||  ||||||
Db      10717 ACGGCGCCGAACGGCCCCGCCCAGCAGCGGGTGATCCGGCAGGCGCTGGCGAGTGCGGGG 10776

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||||  | |||||||||||| | |||||||| |||||| ||||| ||||||||||| |||
Db      10777 CTGTCGGCGGCGGATGTGGACGCGGTGGAGGCGCACGGTACGGGTACGACGCTGGGCGAT 10836

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGG---GCAGCGGGCCGGTGACAGGCCG 1017
              |||||||||||| |||||| |||||||||||||| ||    | ||||  |  |  ||| |||||
Db      10837 CCGATCGAGGCCCAGGCGCTGCTCGCCACGTATGGCCAGGAGCGACCTGAGGACCGGCCG 10896

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077
               ||  |||||||||  | || |||||||||||| ||  |   ||| ||  |||||||||

Db      10897 TTGCTGCTGGGGTCGGTCAAATCCAACATCGGTCATGCGCAGGCGGCTTCGGGTGTGGCG 10956

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137
              || |||||||||||||| ||||| ||||| | || ||| |||| |||||| ||||||||
Db      10957 GGTGTCATCAAGATGGTGCTGGCGATGCGGCACGGTGTGCTGCCTCGGACGCTGCATGTG 11016

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197
              ||| | |||||| ||||  || ||||||    || || || ||  ||||||||||     | |
Db      11017 GATGAACCGTCGTCGCATGTCGACTGGAGTGCCGGTGCCGTCGAGCTGCTGACCTCCGAG 11076

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257
              |    |||||||  || || ||       ||  ||| || ||||| || || || |||||
Db      11077 GCCGAGTGGCCGCAGGGCGAGG------GGCCGCGCCGCGCGGGCGTCTCCTCCTTCGGC 11130

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGC 1295
              || |||| ||||| |||||||||||  ||||| || |
Db      11131 GTCAGCGGGACGAACGCGCATGTGATCCTGGAGCAGCC 11168
```

```
RESULT 10
AX067996
LOCUS       AX067996               20394 bp    DNA     linear   PAT 19-JAN-2001
DEFINITION  Sequence 3 from Patent WO0077222.
ACCESSION   AX067996
VERSION     AX067996.1  GI:12329806
KEYWORDS    .
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 20394)
  AUTHORS   Martin,J.F., Aparicio,J.F. and Colina,A.J.
  TITLE     Genes encoding enzymes in the biosynthesis of pimaricin and the
            application thereof
  JOURNAL   Patent: WO 0077222-A 3 21-DEC-2000;
            DSM N.V. (NL)
FEATURES             Location/Qualifiers
     source          1. .20394
                     /organism="Streptomyces natalensis"
                     /db_xref="taxon:68242"
     CDS             1. .20394
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC24642.1"
                     /db_xref="GI:12329807"
                     /translation="MSNEEKLREYLKRAIADLHETRQQLDETEAKQREPLAIVSMACR
                     FPGGVRSPEELWELLRDGVDAVSSFPRNRGWDLDALYHSDPAHQGTSYAREGGFLHDA
                     GEFDPGFFGISPREALAMDPQQRLLLETAWEAVERAGIDPESLAGSRTGVFVGTGHGG
                     YDAEGRRRADEVGGHLLTGNHISIASGRISYVLGLEGPALTVDTACSSSLVALHLAMH
                     ALRRDECAMALVGGATVMSTPQMFVEFSRQRGLAPDGRCKPFAAAADGTGWSEGVGLL
                     LVERLSDAVRNGYPVLAVLKGSAVNQDGASNGLTAPNGPSQQRVIRQALTGAGLAASD
                     IDAVEAHGTGTTLGDPVEAHALLATYGQQRAADRPCGLGSMKSNIGHTQAAAGIAGVM
                     KMVLAMRHGHLPRTLHLDEPTGHVDWSEGNARLLAEPEPWPSAGRPRRAAVSSFGISG
                     TNAHVILEQAPAHEAEPAPEPAARPGALPWILSARTEAGLRAQADRLGRHLRDRADLE
                     PAAVAHALADTRTLMEHRAVVVAGDREEFLRGLDALAAGRTANGLVSGVAVKAASAFL
                     FAGQGSQRPGMGRELHAAHPVFATAFDAVCAELDPHLDRPLRDIVFAEEGSAEAALLD
                     QTAYTQAALFALETALFRLVESWGVAPRFVAGHSIGELTAAHVSGVLTLHDAARLVAA
                     RGTLMQALPAGGAMVAVQATEDEIRERLAGHEDHVALAAANGPDSTVISGDEQAVTEI
                     AAHWEAQGRRTKRLRVSHAFHSPHMDDMLEDFRRVARGLTFHAPRIPVVSTVTGALAT
                     EDELRSPDYWVRQVRETVRFCAAVRTLEAEGVTTFVEIGTGGVLTPMVQDCLTTLEEP
                     VLVPLLRTGRPETVALTEGVATAFVHGVPVDRSAFPGAPGTSRADLPTYAFQRQWYWL
                     DPADHDEGEAAAAEAGEAGFWAAVEREDLQELSAVLAIDGSEADSLGSFLPTLSSWRR
                     QRRTQAAADRFSYRTHWAPRTASGGPTATGHWLVVLPEGGTDDPWTARLLDALNDQGL
                     HTDVRELPADHEPDAWGRHPVDGVLCLLALDERPTRSCPPYRRGLAATTNAAARPEGA
                     GIQAPLWCVTRGAVAVDRHEALKSPLQAQTWGLGRVAALESPQSWGGLIDLPDNLDGR
                     AVSALLSTLAGEEDQVAVRPAGVFARRLERITPGGDTGDRWSTHGTVLVTGGTGALGA
                     HLAHWLADAGAEHLVLTGRRGPQAPGAPELAAALTDRGVKVTLAACDAADRDALAAVL
                     ADIPPHLPLTGVVHAAGVLDDGVLDALTPERFETVLRPKARAAQNLHELTQDLDLDHF
                     VLFSSIVGVLGNAGQANYAAANAYLDALAEHRLAQGLPATSVSWGPGQAAAWHDSDAA
                     DRMSRDGLLPMAAAPRRRPAPALAQGMTQVTVADIDWSAYAPALTAVRPSPLIGDLPE
                     ARRALGPAEGPRRERSPLRDRIGALPPAEQEKAFLTMVREEAARVLGHPSPDTVDAQR
                     AFREQGFDSLMAVDLRNRLSAATGLRLPATLLFDHPTPLAAAACLRSEVLGAAGPATV
```

macrolide pimaricin. A new polyketide synthase organization encoded by two subclusters separated by functionalization genes
JOURNAL J. Biol. Chem. 274 (15), 10133-10139 (1999)
MEDLINE 99214571
PUBMED 10187796
REFERENCE 2 (bases 1 to 20394)
AUTHORS Aparicio,J.F.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-1999) Aparicio J.F., Microbiology, INBIOTEC, Av. Real, 1-Leon, 24006, SPAIN
FEATURES Location/Qualifiers
source 1. .20394
/organism="Streptomyces natalensis"
/db_xref="taxon:68242"
gene 1. .20394
/gene="pimS1"
CDS 1. .20394
/gene="pimS1"
/function="pimaricin biosynthesis"
/codon_start=1
/transl_table=11
/product="polyketide synthase"
/protein_id="CAB41041.1"
/db_xref="GI:4678703"
/translation="MSNEEKLREYLKRAIADLHETRQQLDETEAKQREPLAIVSMACR
FPGGVRSPEELWELLRDGVDAVSSFPRNRGWDLDALYHSDPAHQGTSYAREGGFLHDA
GEFDPGFFGISPREALAMDPQQRLLLETAWEAVERAGIDPESLAGSRTGVFVGTGHGG
YDAEGRRRADEVGGHLLTGNHISIASGRISYVLGLEGPALTVDTACSSSLVALHLAMH
ALRRDECAMALVGGATVMSTPQMFVEFSRQRGLAPDGRCKPFAAAADGTGWSEGVGLL
LVERLSDAVRNGYPVLAVLKGSAVNQDGASNGLTAPNGPSQQRVIRQALTGAGLAASD
IDAVEAHGTGTTLGDPVEAHALLATYGQQRAADRPCGLGSMKSNIGHTQAAAGIAGVM
KMVLAMRHGHLPRTLHLDEPTGHVDWSEGNARLLAEPEPWPSAGRPRRAAVSSFGISG
TNAHVILEQAPAHEAEPAPEPAARPGALPWILSARTEAGLRAQADRLGRHLRDRADLE
PAAVAHALADTRTLMEHRAVVVAGDREEFLRGLDALAAGRTANGLVSGVAVKAASAFL
FAGQGSQRPGMGRELHAAHPVFATAFDAVCAELDPHLDRPLRDIVFAEEGSAEAALLD
QTAYTQAALFALETALFRLVESWGVAPRFVAGHSIGELTAAHVSGVLTLHDAARLVAA
RGTLMQALPAGGAMVAVQATEDEIRERLAGHEDHVALAAANGPDSTVISGDEQAVTEI
AAHWEAQGRRTKRLRVSHAFHSPHMDDMLEDFRRVARGLTFHAPRIPVVSTVTGALAT
EDELRSPDYWVRQVRETVRFCAAVRTLEAEGVTTFVEIGTGGVLTPMVQDCLTTLEEP
VLVPLLRTGRPETVALTEGVATAFVHGVPVDRSAFPGAPGTSRADLPTYAFQRQWYWL
DPADHDEGEAAAAEAGEAGFWAAVEREDLQELSAVLAIDGSEADSLGSFLPTLSSWRR
QRRTQAAADRFSYRTHWAPRTASGGPTATGHWLVVLPEGGTDDPWTARLLDALNDQGL
HTDVRELPADHEPDAWGRHPVDGVLCLLALDERPTRSCPPYRRGLAATTNAAARPEGA
GIQAPLWCVTRGAVAVDRHEALKSPLQAQTWGLGRVAALESPQSWGGLIDLPDNLDGR
AVSALLSTLAGEEDQVAVRPAGVFARRLERITPGGDTGDRWSTHGTVLVTGGTGALGA
HLAHWLADAGAEHLVLTGRRGPQAPGAPELAAALTDRGVKVTLAACDAADRDALAAVL
ADIPPHLPLTGVVHAAGVLDDGVLDALTPERFETVLRPKARAAQNLHELTQDLDLDHF
VLFSSIVGVLGNAGQANYAAANAYLDALAEHRLAQGLPATSVSWGPGQAAAWHDSDAA
DRMSRDGLLPMAAAPRRRPAPALAQGMTQVTVADIDWSAYAPALTAVRPSPLIGDLPE
ARRALGPAEGPRRERSPLRDRIGALPPAEQEKAFLTMVREEAARVLGHPSPDTVDAQR
AFREQGFDSLMAVDLRNRLSAATGLRLPATLLFDHPTPLAAAACLRSEVLGAAGPATV
VQASTAALDEPVAIIGMACRFPGGVHSPEALWRLLAEGGDAITPMPADRGWDLDRLYH
PDPDHQGTSYARGGGFLDGAADFDADFFGISPREALAMDPQQRLLLETWEVLEQAGID
PESLRGSSTGVFAGTNTQDYGTALDAAQDEAGGHRLTGNAMSVVSGRVSYTFGFEGPA
LTVDTACSSSLVALHMAAQALRQGECSLAVAGGVTVMATPSSFVEFARQRGLAPDGRC
KPFAAAADGTGWSEGVGLLLVERLSDARRNGHQVLAVVRGSAVNQDGASNGLSAPSGP
SQQRVIRQALANARVAASEVDAVEAHGTGTTLGDPIEAQALLATYGQERPLLLGAVKS
NLGHTQAAAGVAGVMKMVLAMRHGMLPRTLHVDEPTGHVDWTAGAVELLTEHTDWPET
GHPRRAAVSAFGISGTNAHVVLELPAAEQPLVEQPSAAEPDAPATAPDRTPTASDGTA
PLLLSAKSESALRAQAARLHSHLERDPALRLTDAAYTLMTHRTAFAHRAAVRAADHEA
ALRALTALAAGEADPAVDTGTAHTGRDAVLFSGQGSQRIGMGRELSGRYPVFAEAFDT
VCAALDEHLDRPLRDVVRGEDEELLNRTVYAQAGLFAIEVALFRLVESWGVRPHYVAG
HSVGEIAAAHVAGVFSLADACALVAARGRLMQALPAGGAMAAIRATEDEVLPHLADSV
SIAAVNGPSSVVVSGAEHAVLSIAAHFEGAGRKTTRLRVSHAFHSPLMDPMLADFRAV
AEGLTYGEPELAVVSNVTGQLATPDQLRTPEYWVTHVRAAVRFADGIRALGAEGVTRF
LELGPDGVLSALARESAPDDAVCTPVLRKDRSEAATLLAALTHLHVHGTEIDWTAFLA
GRDAHAVDLPTYAFQHQRFWPTPDHTRTGDLGAVGLEATGHPLLSAAVELPDGEGLLF
TTRLSLQTHPWLAGHVVMGSVLLPGTAFAELALRAADEVGCDRVDELTLAAPLVLPEH
GGVQLQLRVGPADASGRRTLTARSRAEGDGDRPWVQHATGVLAEGESTPEPGYDFHTE
SWPPADAAPVELSGLYPDFAAHGFDYGPHFQGLRTAWRRGDEVFAEVALPAEAEGEAS
AYGLHPALLDAALHVVAFNGVDRGVVPFSWESVALHATGASAVRIRVVRHSGDTVSVD
VADTTGEPVASIGTLVLRAVSADQLAGGADPAVRDALFRVQWNPVRLPPAGAAVTVAT
LGSLAGAPFDGYPDLASLARSGRVAGAVLVPVEAGAGEVVADDVVGATHATAARALDL
ARSWLADDRFAASRLVFVTRGAVSGADLAGAAVWGLVRSALSEHPGRFGLVDLDDDAE
LALVPRVLASDEPQLLVRGGEVLAARLARAQSSHAVTWDPSGTVLVTGGTGGLGRVMA
RHLVVEHGVRNLLLVSRRGPAAEGAEELVTELRHSGAEVAVEACDVTDAAAVADLVAR
HRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRDLDLDAPVVFSSV
AGTIGSPGQANYAAGNAFLDALAHHRRAAGLPAASLAWGPWSRDGGMTGTLTDVDSSA
SPGRHARTHPRTGRGLFDAALAAGDAHLLPVRFDWASLRAQGEVPPLLRGLIRTRARR
SAVGGSAAAAGLVGRLSGRGTVERREVLLDLVRAQIAVVLGHANPETIESTRVFQDLG
FDSLTAVELRNRLNNATGLRLSATAVFDYPTADALVDFLLDELFGAQEEAELPAPVPS
PAGAADDPVVIVGMSCRYPGGVGSPEDLWRLVSEGVDAVSDFPTDRGWDVESLYSPDP
EALGTSYTRSGGFLHEAAEFDPDFFGMSPREALATDAQQRLLLETTWEAIERTGIDPA
SLRGSRTGVFAGVMYTDYGDLLVGDQFEGYRSNGSAASIASGRVSYTFGFEGPAVTVD
TACSSSLVALHWAAQSLRSGECSLAVAGGVTVMSTPTTFVEFSRQRGLSADGRCKAPA
DAADGVGWGEGVGMLVLERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGPAQQR
VIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLATYGQERPEDRPLLLGSVKS
NIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHVDWSAGAVELLTSEAEWPQG
EGPRRAGVSSFGVSGTNAHVILEQPGPDAADAAPDATVTDPGALAWVLSARNEAALRC
QAARLLSLVAGSDALCARDIGHSLVTGRSSFAHRAVVWGQDRDALVRALSALAVGEAD
AGLAEGASGAGRTAFLFSGQGSQRLGMGWELYARYPVFADAFDAVCAALDEHLERPLR
DVVWGEDAELLNQTAYAQAGLFAIEVALYRLAESWGMRPDFVAGHSIGEVAAAHVSGV
FSLPDACALVAARGRLMQQLPSGGAMMAIRATEDEVLPHLAEGVSLAAVNGPSSVVIS
GAEDAVLAIAAHFAGEGRKTTRLRVSHAFHSPLMEPMLEEFRAVVTRLSFGTPTIPVV
SNLTGRLAEPEQLAHADYWVRHVREAVRFADGIQALRAEGVTRFLELGPDGVLSAMAR
ESASDDAVLAPVLRRDRPEETALLGALAQLYVRGAHVDWTVPFAGSGARWADLPTYAF
QHERFWPSGGVARPGDVRSAGLGSAGHPLLGAAVELAGSGGLLFTGRLSVSSHPWLAD
HVVLGSVLVPGTALVELVLRAADEAGCDLLEELTLAAPLVLPASGAAVQVQVAVGEPD
EAGRRPVSVHAREGEGPWTLHASGAVTSGAEVPPFDATVWPPKGAEPVDVADCYDVLA
DAGLTYGPAFHGLQAAWKLGGDVYAEAKLPESTDGDAYGLHPALFDAALHASALGGAE
AGGVPFSWAGVSLHATGASHLRVRIREAGGALSVAIADTSGAPVASVESLVIRPLSAG
QVQAADRDALFKADWVPVPLTDERVEPGTGPEGEPLRTYADLDSLEGAAVPGTVLVAP
PSGAAGTVESVHAATVWALEMVQAWLADDRFATSRLVFVTRGAAFGADLAAAAVRGLV
RSAQSENPGRFGLVDMDGDADTTVPAQALATDEPELLVRGGEVLAARLVRAQSSHTVT
WDPSGTVLITGGTGGLGRSVARHLVSEHGVRSLLLVSRRGPAAEGAGELVAELRGSGA
EVVIEACDVTDAVAVADLVARHRISAVVHTAGVLDDGVVESLTPERLAVVLRPKVDAA
WNLHEATRGLDLDAFVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLA
WGPWSQDGGMTGTLSDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLA
ALRAQGEPQPLLRGLIRTRTRRSGAAAASGIAQRLAGLSTAERREALLDVVRAQIATV
LGHAGPETIAPDRAFQDLGLDSLTAIELRNLLGKATGLRLPATTVFDYPTVDALAAHL
LDELFGAETGTATETPLPVPGLPSLADDPVVIVGMSCRFPGGVASPEDLWRLVADGVD
AVSAFPTDRGWEIDDTYDPEREGAIATRSGGFLHDAAEFDPEFFGMSPREALTTDAQQ
RLLLETTWEALERAGMDPATLRGSRTGVFAGVMYHDYSTLLSGREFEGYQGSGSAGSV
ASGRVSYTFGFEGPAVTVDTACSSSLVALHLAAQSLRSGECSLALAGGVTVMSTPLTF
VEFSRQGGLSADGRCKAFADAADGVGWAEGAGILVLERLSDARRNGHRILATVRGSAV
NQDGASNGLTAPNGPAQQRVIRQALASAGLSAADVDAVEAHGTGTTLGDPIEAQALLA
TYGQERPEDRPLLLGSVKSNIGHAQAASGVAGVIKMVLAMRHGVLPRTLHVDEPSSHV
DWSAGAVELLTSEAEWPQGEGPRRAGVSSFGISGTNAHVILEQPEPVAAETESITPDT
APDAAEDEAADSGTPVPALLSGRSASALRAQAARLLSRLDGDPGPRITDVAYSLATGR
SAFPHRAVILAANRADLLHSLSALAEGHTEAPAVVAQDRARSGKLAFLFSGQGSQRLG
MGRELYGRYPAFAEALDAVCAALDAHLDRPLRDVIWGEDAELLNRTGYAQTGLFAIEV
ALFRLLESWGVRPDHLLGHSIGEIAAAHVAGVLSLPDACALVAARGRLMQQLPSGGAM
MAIRATEDEVLPHLAEGVSLAAVNGPSSVVVSGAEDEVLALAAHFEEEGRKTTRLRVS
HAFHSPLMEPMLADFRAVADGMTYAAPRIPVVSNVTGRPATAEELCCAEYWVGHVREA
VRFADGVGALREQGVTTFLELGPDGSLSALAAESAADDSVLAPVLRKNRPEAPALLTA
LARLHAQGTPVDWSAAFAGTGARWVDLPTYAFQHERFWPSGGAARAGDVRSAGLGSAG
HPLLGAAVELAGSGGRLLTGRLSLSSHPWLADHVVLGSVLVPGTALMELVLRAADEVD
CAAVDELTLAAPLVLPASGAAIQVQVWVGEPDEAGRRPVSVHAREGEGPWTLHADGAL
APAAETVPFDTAIWPPQGAEHLDAAGCYERFADAGFAYGPVFQGLRAAWKLGEDIYAE
VALPEGTDGNAYGLHPALFDAALHAALLGGEGTDEAAVPFSWNGVTLHATGASRVRVR
IRPTEGGTSIALVDTAGAPVASVRSLTARPITAGQLQTGDRDSLFQVDWTTLHLTDER
ANSLALLGKDTEGILDTLSLQPHADLDDLAATGVHDTVLAPLPTRTAGTVESVHAATT
GALALIRSWLADDRFAASRLVFVTRGAVSGTDLAGASVWGLVRSALLEHPGRFGLVDV
DVDQDAEVPLVPRALASDEPQVLVRGGEVLAARLVRAQSSDTVTWDPSGTVLITGGTG
GLGRSVARHLVSEHGVRSLLLVSRRGPAAEGVDALVAELAECGAQVTVEACDVTDAVA
VADLVARHRISAVVHTAGVLDDGVVESLTPERLSAVLRPKVDAAWNLHEATRGLDLDA
FVVFSSVAGTFGSAGQANYAAGNAFLDALAYHRRAVGLPAVSLAWGPWSQDGGMTGTL
SDADVQRIARQGMPPLTVEEGLALFDAALGSAEPMALPVRLDLAALRAQGEPQPLLRG
LIRTPGRRTAAAATEGDTAAAFAGRLTGLSAAEGREVVLGAVRSQIAGVLGHAEATEI
DQDRAFLDLGFDSLTAVELRNRLGAVTGIRLPATLLFDYPTPAELVAHLHARIAPEPT
VGPEALLGELERMEKSFGGLDLTEEMHEQIAGRLEVLRAKWDALRDTAAAAGHDGSPS
DEDFDFESASDDEVFDLLDNELGLS"
BASE COUNT 2611 a 7486 c 7397 g 2900 t
ORIGIN

Query Match 54.5%; Score 710; DB 1; Length 20394;

```
                     /db_xref="GI:21999186"
                     /translation="MGRQVCAVGSLRREEGGLRRFLTSAAEAFVQGVDVTWPALFDDT
                     GARAIDLPTYPFQHQNYWYIDPARRKGDVSSFGMAEAGHPLLDAGTELPEGGEHLYTA
                     RLTADSHPWLLEHTLLGTPLLPGAAFVDLVLWAGGELGCELVEELTLTSPLLLSDSAA
                     VQLRLVVGATDAEGRRTITIHSRPDAHPGTTRTPAAPPEAGSEAEVRRDAADWTKHAQ
                     ATVAPAPAAPASVVGADGNPGLPAAEWSSAATESDTFRAEDFYASFAARGYGYGPLFQ
                     GVRSGRQDGSDVYAEVALDQGALPPGERFGLHPALLDAAFQTMRLGSFFPDDGQARVP
                     YTFRGIRLYAPGAERLRVRVSAVGDDAVRVECADEQGRLVCEIDALVVSPVSPAQLRP
                     AGQDATQDMLHRVEWPVLPSPAGRATSTAAFRWIVVGPEDDGLGLAEVLSGTGTGIER
                     HRNLADALTALRTPADTPGSPAAADKTPLIAVPVSPSPTMDAVSVRHAVHQALESVQG
                     WLSADEAAGEGGSGAAGADGRLVLITSGAVSTGDADPLRDPVASAVWGLVRSAQSEQP
                     DRFVLVDLDEGAVDGAALSAAIASGEPQLALREGKVHVPRLASLSARDSRMLLPPAGA
                     RAWHLVGARTGTLSDLALVPAQADTVPLAPGQVRIAVRAAGLNFRDTLIALGMYPGEG
                     VMGAEGAGVITEVGPDVAGFAVGDRVLGMWTDGFGPYAVADHRMVAPMPRDWSYAEAA
                     SVPAVFLSAYYGLKHLAGLRAGQSVLVHAAAGGVGMAAVQLARYFGAEVFGTAGPAKW
                     DVLRAQGLDDRHIASSRTLDFAARFLDATGGRGVDVVLNSLAGDFVDASLRLLPRGGR
                     FVELGKADVRDAAQVAADRPGTGYRAFELMEAGPELIGRMLTELLELFESGALRLLPV
                     TPYDIRRAPDAFRTLSQAGHVGKLVLTMPPAFDPPRPGRPACSCRPRHRTRRRGTRSA
                     ARPGAPARRCRTLVRCGHAGPCGGPGAPDHGRGIRHPGRTARREAARRTARRPQGARA
                     HRNRLGARARGRPTRPTPVLLQGSRIRLAHRRRTPQQAHRRHRNEASRHTGLRPSDAG
                     GTRRPPGTGTAEGRSETSRGRPGRSGGTRSDPGRTGHRCHRRQRPRPRRGTPSGTARH
                     ARPAGGGRPPGQRLERSRRAGPR"
     CDS             36783. .37874
                     /note="Orf5"
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAM81588.1"
                     /db_xref="GI:21999187"
                     /translation="MNDEWLIYRGVGEPHDGIGALPGPPPWRDFDGGPAAEPADAAAG
                     EGGGAAGAGGGQGPDAPDGNVARRLGAHRQAAEAHRPEPEELEAVNAALYLRRPLLVT
                     GIPGTGKSTLAHAVAHELKLGRVLRWPVVSRTVLQEGLYRYDAIARLQDVQIAAGGRA
                     GGAAGGASGGASGGSSGIGRYIRLGPLGTALLPTARPRVLLIDELDKSDIDLPNDLLN
                     VLEEGEFTLPELERVADTEPEVRVLTDDGAKVTVRDGRVRCRAFPFIVLTSNGERDFP
                     AALLRRCIQLKLGQPGEKRLATMVRAHLGEEAARLGADLIREFLSRSRTELVAADQLL
                     NAIYLTHYAAPPTREDLADLLIQRLDRPR"
BASE COUNT     5228 a  14409 c  16025 g   6119 t      1 others
ORIGIN

  Query Match             55.5%;  Score 722.2;  DB 1;  Length 41782;
  Best Local Similarity   73.3%;  Pred. No. 2e-82;
  Matches  954;  Conservative    0;  Mismatches  338;  Indels    9;  Gaps    2;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || ||||| || ||| | || |||| |||||  ||||||| || ||   ||| ||| ||
Db      21885 GAGCCGATTGCGATCGTGGGGATGGGGTGCCGGCTCCCGGGTGGGGTGTCGTCGCCGGAG 21944

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              |   ||||||| |||| || ||||| |  |||||  ||  |||||||| ||||| ||
Db      21945 GGGTTGTGGGATCTGGTGGCGTCGGGGGTGGACGCGGTCTCTCCGTTCCCTACCGATCGG 22004

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||     ||      |      | . | ||| ||       ||| ||        ||| |
Db      22005 GGTTGGGATGTGGCGGGTCTGTTTGATCCGGAGCCGGGTGTGCCTGGGCGTACGTATGTG 22064

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              |  |  || |||||||| || |  |||| | | |||||||| || || ||||| ||
Db      22065 CGTGAGGGTGGGTTCCTGCATGGGGCGGGGGAGTTCGACGCGGGGTTTTTCGGTATTTCG 22124

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              || ||||||||  |||||||||||||||||||||| || ||||||||||||| ||||||
Db      22125 CCGCGTGAGGCGTTGGCGATGGATCCGCAGCAGCGGTTGTTGCTGGAGACGTCGTGGGAG 22184

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||||| |||||||||| |||||||||| ||| | ||||| || || ||||||| |
Db      22185 GCGTTGGAGCGGGCGGGTATCGATCCGCATTCGCTGCGCGGCTCACGGACCGGCGTCTAC 22244

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              || || |   | | |  |||||||||        |||   | |  ||  | |    || |
Db      22245 GCCGGAGTGATGTACCACGACTACGGCAGCACCGTGG---CGGGCTCCGCCGCTCCCGAC 22301

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              |  ||   ||  |||  | ||  || ||| ||  | ||||| |||   ||||| ||||||

Db      22302 GACGAGACGGCCGGCTTCCTCGGCATCGGCACGTCCGGCAGTGTCGCTTCGGGTCGTATC 22361

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              | ||     | || ||||| |||||||| | || |||||||||| ||| ||| | |||
Db      22362 TCGTATGTGCTGGGTCTGGAGGGGCCGGCGGTGACGGTGGACACGGCGTGTTCGTCGTCG 22421

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              || ||  || |||||||||||    | | |  ||||    ||||||||    ||||| ||
Db      22422 CTGGTTGCGTTGCATCTGGCGGTGCGGGCTTTGCGCAGTGGTGAGTGTGATCTCGCTCTT 22481

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              || || || ||  |||||||| |      |  ||  ||||| | |||||||| ||||||||
Db      22482 GCGGGTGGGGTGACGGTCATGGCGGAGCCGGGTGTGTTCGTGGAGTTCTCGCGGCAGCGG 22541

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              || ||| | |  || ||  |||||||||||| |||||    | ||||||||||| ||||||
Db      22542 GGTCTGGCCGCCGATGGGCGGTGCAAGGCGTTCTCGGACTCGGCCGACGGCACGGGCTGG 22601

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              | |||||||| || ||||  |||||||||||||||||||||||||||| |  || |||
Db      22602 GCCGAGGGCGTGGGTGTGCTGGCGGTGGAGCGGTTGTCGGATGCGGTGCGTCATGGACAT 22661

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              |||||||||||||||| || ||     ||||| || ||||||||||||   |||||| ||
Db      22662 CGGGTGCTGGCGGTGGTGCGTGGTTCGGCGGTGAATCAGGACGGTGCGAGCAATGGGTTG 22721

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              |||||||||| |||||| | ||| |||||||||||||||||||||||| | ||| ||
Db      22722 ACGGCGCCGAATGGTCCGTCGCAGCAGCGGGTGATCCGGCAGGCGTTGGCGGATGCGCGG 22781

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              |||   ||||  |||||||||||||||||| ||| |||||||| |||   |||||||||
Db      22782 TTGGGTGTGGGTGATGTGGATGTGGTGGAGGCGCATGGGACGGGGACGCGTCTGGGTGAT 22841

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              |||||||| || ||||||||| | || ||||| |||||||| | ||  |   ||||  ||
Db      22842 CCGATCGAAGCGCAGGCGTTGTTGGCGACGTATGGGCAGCGCGACGCGGGGCGGCCTTTG 22901

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              ||||| || || |||||||| ||  | || || ||   |||||| |||||||||| |||
Db      22902 TGGCTTGGTTCGCTGAAGTCGAATGTGGGTCATACGCAGGCTGCTGCGGGTGTGGCTGGT 22961

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              || |||||||||||||||||| ||||||| || ||  ||||  |||| ||||||||||
Db      22962 GTGATCAAGATGGTGATGGCGATGCGGGATGGAGTCCTGCCGAAGACGCTGCATGTGGAT 23021

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              |||||||| || |||||||||||| ||||||||||    |||||||||| || |
Db      23022 GAGCCGTCGGCGGAGGTGGACTGGTCGGCGGGGGCGGTGTCCTTGCTGACGGAGCAGGAG 23081

Qy       1201 CCGTGGCCGGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              |||||||||      | ||  |||||   |||  |||| |  ||||| || || ||| ||
Db      23082 CCGTGGCCG------GAGGTGGGGCGTCCGCGCAGGGCCGCGGTGTCCTCCTTTGGGCTC 23135

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGC 1301
              | || ||||||||||||||| | ||||||||||||||| ||
Db      23136 AGTGGGACGAATGCGCATGTGGTGTTGGAGGAGGCGCCCGC 23176


RESULT 9
SNA132222
LOCUS       SNA132222              20394 bp    DNA     linear   BCT 24-APR-1999
DEFINITION  Streptomyces natalensis pimS1 gene.
ACCESSION   AJ132222
VERSION     AJ132222.1  GI:4678702
KEYWORDS    pimaricin biosynthesis; pimS1 gene; polyketide synthase.
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 20394)
  AUTHORS   Aparicio,J.F., Colina,A.J., Ceballos,E. and Martin,J.F.
  TITLE     The biosynthetic gene cluster for the 26-membered ring polyene
```

```
REFERENCE   2  (bases 1 to 41782)
  AUTHORS   Lim,S.K., Nam,D.H., Sohng,J.K. and Yoo,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-2002) Biotechnology, Dae-Dong, Kyungsan,
            Kyungpook 712-749, S. Korea
FEATURES             Location/Qualifiers
     source          1. .41782
                     /organism="Streptomyces sp. GERI155"
                     /strain="GERI155"
                     /db_xref="taxon:201500"
                     /clone="cosmid pGK11"
     CDS             10272. .15671
                     /note="Orf1"
                     /codon_start=1
                     /transl_table=11
                     /product="putative polyketide type 1 synthase"
                     /protein_id="AAM81584.1"
                     /db_xref="GI:21999183"
                     /translation="MDPQQRLLLETSWEALERAGIDPHSLRGSRTGVYAGMFHQEYAA
                     RLHEAPAEFEGHLLTGTSGSVASGRISYVLGLEGPAVTVDTACSSSLVALHLAVRALR
                     SGECDLALAGGVTVMAEPGVFVEFSRQRGLAADGRCKAFSDSADGTGWAEGVGVLAVE
                     RLSDAVRHGRRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALADARLGVGDVDV
                     VEAHGTGTRLGDPIEAQALLATYGQRGAGRPLWLGSLKSNVGHTQAAAGVAGVIKMVM
                     AMRDGVLPKTLHVDEPSAEVDWSAGAVSLLTEQEPWPRGERVRRAGVSAFGVSGTNAH
                     VVLEEAPDVRGSLEPGAAGLLPVVPVVVSGRSAGVVAELASRMGEVAGSGRLVDVGWS
                     SVVSRSVFEHRSVVLAEESVDLSAGLDAVAVGDGLSPGVVSGVVSVEGARSVFVFPGA
                     GAKWSGMAARLWAESGVFAESMARCEAAFEGLVDWRLSDVIAQVPGAVSLEREDVAQP
                     ASFAVMVSLAALWRSLGVVPDAVVGHSQGEIAAAVVAGGLSLEEGARAVVFRSRVAEE
                     VLSGGGIASVRLSRAEVEERLAGGGGGLSVAVVNAPSSTVVAGELGELERFVAACEAE
                     GVRARRLEFGYASHSVFVEPARGRLLEALADVRPVSGEIPFYSTVEAAVIDTASLDAE
                     YWFGNLRRPVRFQETVEQLLADGFRVFVECGAHPVLTAAVQETAESVGRQVCAVGSLR
                     REDGGLRRFLTSVAEAFVQGVDVTWPALFDGTGARTVDLPTYPFQRQHYWAQASPAGT
                     GSSAAARFGMTWEEHPLLSGALPLAESGELLLVGRISPASHSWIADHTVAGTALLPGT
                     AFVGMALHAAAVVGCAGVEELSIEAPLPVHGGIRLQVVIDEPDSSDRRRVSVFARPEA
                     EDGDTGRWTRHATGVLTPAAAPDPGRPQWSREAWPPSGSVRAEAAELYDRFSALGYEY
                     GEVFAGVEAVWLREGEVFAEVRLPAGAAPDAERFGVHPGLLDAALQPWLLGDFLSRPD
                     DGSVLLPFAWRGITLHATGADALRVRLGPAGEGAMSLEAVDLAGAPVLSMDALVLRPL
                     AQDRLVELVGGTTSAPLYRVDWQRSPIANTAPSATGLFGSFPSGATRRWAVVGQGGAA
                     ARYATADPGTGCVGVFPDLDALRTALDSGADRPDIVLADFGARPGTAAPHGTALADGV
                     RDTARRGLALVQGWLSDERFAAARLVVLTEHAVATEADERQVDLAGSALWGLMRSAQT
                     EHPDRFVLVEHDGQDASYQALPAALDSGIPQLALRSGKVRTPGLAVLPTPGDGLTATG
                     TGFDPEGTVLVTGATGTLGSLLARHLVTAHGVRHLLLLSRSGRDAGRAAELERELTGL
                     GAEFQLLSCDATDRVALQEALATVPAAHPLTAVIHTAGVLDDGVVEALTPERLDRVLR
                     PKADAALNLHELTAGMPLKAFVLYSGAVGLLGGPGQANYAAANAFLDGLAQYRHAQGL
                     PAVSLAWGLWSAASTFTGHLGEVDLRRMERSGIAPLTNEQGLELLDRALGAAAGADAP
                     QICAMRLDTAALRRQASEHGPASVPLLLRTLAAPAVRRGAARTGRGARAAGAVTDVPS
                     RAQDLRERLAGLDAAGRRDELLAVAQAQLADVLGFTDKTAVDPVRSFREIGLDSLTAV
                     ELRNRLDAVTGLRLSSALVFDHPNLDALAGHLAELLMAEGKDDAGAAALSGVDALDRA
                     VRGMAADDTRRDAVRRRLTELLAVVGDAPRDGDRSTRTAAGKGSHGVQAGPDLLGQLD
                     SASDDDLFAFIEDQL"
     CDS             15768. .21698
                     /note="Orf2"
                     /codon_start=1
                     /transl_table=11
                     /product="putative polyketide type 1 synthase"
                     /protein_id="AAM81585.1"
                     /db_xref="GI:21999184"
                     /translation="MTANDDKIRDYLKRVVAELHSTRQRLNALEDNAREPIAIVGMSC
                     RLPGGVTTPESLWELVDSGTDAVSPFPDDRGWDLDALHHPESGAVHSREGGFLHDSAD
                     FDAEFFGISPREALAMDPQQRLLLETAWEVFERAGIDPASVRGSRTGVYAGVMYHDYG
                     ARLNEIPPGLEGYLVNGSAGSIASGRVAYTLGLEGPAVTVDTACSSSLVAVHLAAQAL
                     RQRECDMALAGGATVLSTPDLFTDFARLGGLASDGRCKAFSDAADGTSFAEGAGMLLL
                     MRLSDAVDKGHTVLAVIRGSAVNQDGASNGLTAPNGLAQQRVIREALANADLGPDQID
                     AVEAHGTGTRLGDPIEAQALLHTYGTSRSPERPLLLGSLKSNIGHTQAAAGVAGIIKT
                     VLAMRHGRLPRTLHVTRPSSRVDWSAGAVELLTQTQDWPGQGDAPRRAGVSSFGASGT
                     NAHLILEGVPDSDDTAAGREPSAGGGAWLLSGRTEAALRAQARRLHGHLAARPHVSPA
                     AVGRTLARSRTPFEHRAVVIGQDTAALLSGLTELASGAAHGPGVITGRAARGRRTALL
                     FTGQGSQRPGAGRRLYERYEVFARALDETAAELDRHLDRPLREVMFTEPGSATAGLLD
                     RTEYTQPALFALEVALFELVSAWGLRPDALLGHSVGELAAAHVAGVFTLPDAARLVTA
                     RGRLMGELPAGGAMAAIQAGGPEIEETITALTARRPARIAVAALNGPDATVISGDEDA
                     VAELAALWRERGRRAKVLPVSHAFHSPRMDAALEPFARIARDVSYAEPRIPVVSNLTG
                     EVASAATLCTPEYWVRHAREAVRFSDGFRTLRDQGIDTFVELGPDGVLSALGRDCLRE
                     EEQDAPHQDGSTDRDTAGSPADGQERPVLTVPLLRRDRDEATTCLGALAAVHTHGVPV
                     DLAAVHGDGPAVDLPTYAFQRTRYWLDVPAPAAGLTTTGLEAADQPLLPAAIDLPDGE
                     GTVRTGLLSVRTHPWIADHRVQNRTVVPGAALLDIAAWAGTEAGCPRVAELTFATPLV
                     LPESGEGVRLRVTLSGPDAEGIRTIRIDSRPADAARAADAPSDWTRHASGTLTPGTEE
                     AGDGADVPADLLGVWPPADATPVALDADAVAAEYRRLADAGVTYGPVFRGLRAVWRRG
                     EEIFAEVRLPGQAAADAPRYGMHPALLDALAHATGFGEQFTEAHGLLPFSWSDVRIHV
                     RGTDSLRVRIAPAGPDAVTVAAADPAGRPALTARSLTLRPVPAHRFPSPQERSTPLYR
                     LEWTSASAPVTGHARPRPAAAEWGVLGDAGPALLKGVRDGAQAPVRTYDDLAALAASD
                     TPLPDHVLVELGHDGDDLAAGAHDLAERALAVVQGWLAHARFTDARLVVLTHGAVAAG
                     TSAVRPASAVAWGLLRSAQSEHPGRFVLVDADPSDPAASYRSLSRAVTSGASQLALRG
                     DEILVPRLTHGANGQGAAAGPLGDATEPEAAAAMGSAPSGSLSGPWRADGTVLVTGGT
                     GTLGKAVARHLVTKHGVRHLILAGRRGADAPGAAGLAAELTDLGAVVNIVRCDAADRS
                     ALEGLLAAVPAAHPLTAVVHTAGVLDDGIVTAQTLQRVSAVLRAKADAVTHLHELTRD
                     LDLSAFVLFSSAAGTLGSPGQSGYAAANSFLDAFAAWRRAQGLPAVSLAWGLWGDGGD
                     GGDGDGGSAGADGMGASLAAADLARLRRSGILPLDPAEALELFDEACDPSRTEAVLLP
                     IRLDLAGLRARAARGTVHAGVVPEVLHALVPPPAGAGPSSGAGTPDPAAGQEPPTTAA
                     ASGTLAEQLAGKPRGERLAVLRELVRTEIASVLGHADAQRVQLQFSFKEAGFDSLTAV
                     ELRNRLTAVTGTRLPATLVFDHPTPAALVDHLEQELPKDGQKLPADVPAVLEALDRIR
                     DGLATAVTDDNGRAHVAERLQELLGTLAPPVAGVRPVNGSNGHDGQGPDELSVGERLA
                     ASSDDELFDLFDSDFRSM"
     CDS             21828. .28754
                     /note="Orf3"
                     /codon_start=1
                     /transl_table=11
                     /product="putative polyketide type 1 synthase"
                     /protein_id="AAM81586.1"
                     /db_xref="GI:21999185"
                     /translation="MADLHEAREQISRTESARHEPIAIVGMGCRLPGGVSSPEGLWDL
                     VASGVDAVSPFPTDRGWDVAGLFDPEPGVPGRTYVREGGFLHGAGEFDAGFFGISPRE
                     ALAMDPQQRLLLETSWEALERAGIDPHSLRGSRTGVYAGVMYHDYGSTVAGSAAPDDE
                     TAGFLGIGTSGSVASGRISYVLGLEGPAVTVDTACSSSLVALHLAVRALRSGECDLAL
                     AGGVTVMAEPGVFVEFSRQRGLAADGRCKAFSDSADGTGWAEGVGVLAVERLSDAVRH
                     GHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALADARLGVGDVDVVEAHGTGT
                     RLGDPIEAQALLATYGQRDAGRPLWLGSLKSNVGHTQAAAGVAGVIKMVMAMRDGVLP
                     KTLHVDEPSAEVDWSAGAVSLLTEQEPWPEVGRPRRAAVSSFGLSGTNAHVVLEEAPA
                     PKVGEVGVGVSPAVVPVVVSGRSAGVVRELAGQMGGVAGSGRLVDVGLSSVVSRSVFE
                     HRSVVLAENTTELNTGLDAVAAGVPSSGVVSGVVSGEGGRSVFVFPGQGTQWVGMAVR
                     LWAESGVFAESMARCEAAFEGLVDWRLSQVLDDEAALERVEVLQPASFAVMVSLAALW
                     RSLGVVPDAVVGHSQGEIAAAVVAGGLSLEDGARVVVLRARLIGRELAGHGGMASVAL
                     PVGVVEERLARWAGRLGVAAVNGPSATVVAGDTDAVAEFVTTCEAEGVRARRLPVDYA
                     SHSAHVETLAAELEEVLADVRPVSGEIPFYSTVEAGLVDTASLDAGYWFGNLRRPVRF
                     QETVEQLLADGFPVFVEASAHPVLTAAVQETAESVGRQVCAVGSLRREEGGLRRFLTS
                     VAEAFVQGVDVTWPALFDESGARLVDLPTYPFQHQRYWLESRPSAAVVPSGTQDGLSY
                     GVTWKSLSLPESVRLDGRWLLVVPEHLDADGTRVAHDVEQALTTHGATVHTLALDPAT
                     ADFDGRFSRMFEEATDVTGILSLLGLAVGPHPEHGEVPLGVATSLTLTRQALRAGFGV
                     PVWAVTRGAVSVVPGEVPETAGAQLWALGRVAALELPDRWGGLIDLPADTDARSVGLA
                     VRALTAGIAAGEDQLAVRPSGAYGRRLVRATARRGRKDWRPQGTVLLAGDLDAVAEPL
                     VRWLLTDGADHVVLADPALTELPAALADVAQRVTTAAAPDLADRAVLAALVTEYAPAM
                     VVVVPPAVELAPLESTSPADLAAAVAAKSATAAHFDALLDGPHAPELVLISSVAGIWG
                     GVRQGAYAVGAAHLDALAARRRAQGLPTVSVAWTPWAGSVIADGSAAESMRQYGIAPL
                     EPQTALAELNRALGSGTGDAAVADVDWEQFLASPTSVRPTVLFDELPEVRRLRQAEAA
                     AMADQAARIGAPGGTELARSLRDKPRNAQRKALLELVTAHVAAVLGEGAPDAIDQSRA
                     FKDIGFTSMTAMELRNRLKEATGLTLPASLVFDHPHPRALADHLREELLGEDDAERTD
                     SEPNTPSGPPPAGQDEPIAIIGMACRLPGDVGTPDELWELLETGRDAITDLPVNRGWD
                     VAGLYDPDPDAAGRAYVREGGFLHDAGEFDAGFFGISPREALAMDPQQRIILELAWES
                     FERAGLDPAGRRGSRTGVFMGTNGQHYMPLLQNGSESFDGYLGTGNSASVMSGRISYT
                     LGLEGPALTVDTACSSSLVALHLAVRALRSGECDLALAGGATVMSTPEVLVEFSRQRA
                     VAADGRCKAFSASADGFGPAEGAGVLLVERLSDAVRHGRRVLAVVRGSAVNQDGASNG
                     LTAPNGPSQQRVIRQALADARLGVGDVDVVEAHGTGTRLGDPIEAQALLATYGQRDAG
                     RPLWLGSLKSNVGHTQAAAGVAGVIKMVMAMRHGVLPKTLHVDEASPHVDWSAGAVAL
                     LTEQEPWPEVGRPRRAAVSSFGLSGTNAHVVLEEAPVPEEGQPVGDAPLAVVPVVVSG
                     RSAGVVRELAGQMGGVAGSGRLVDVGLSSVVSRSVFEHRSLVLAEDSAELNTGLDAVA
                     AGVPSSGVVSGVVSGEGGRSVFVFPGQGTQWVGMAVRLWAESGVFAESMARCEAAFEG
                     LVDWRLADVLGDEAALERVEVLQPASFAVMVSLAELWRSLGVVPDAVVGHSQGEIAAA
                     VVAGGLSLEDGARVVVLRARLIGRELAGHGGMASVALPVAVVEERLARSAGRLGVAVV
                     NGPSATVVAGDTDAVAEFVTVCEAEGVRARRLPVDYASHSAHVETLAAELEEVLAGIR
                     PRAGEIPFSRRWKPRRSTRHRWMRATGSGICVGRCGSRRRSSVSSRMVSRCSWRPART
                     RC"
     CDS             28784. .32293
                     /note="Orf4"
                     /codon_start=1
                     /transl_table=11
                     /product="putative polyketide type 1 synthase"
                     /protein_id="AAM81587.1"
```

```
                     LTGASGALGGLVARHLVAEHGVRSLLLLSRRGAEAPGAVELEAELAAWGAEVRWAACD
                     VADREAVSGMLNGLGEQSLSAVVHTAGVLGDGIVASLTPERMREVFRPKVDAVLNLHE
                     CTRDMGLAAFVVFSSVAGMVGSAGQASYAAANSFLDAFSAHRRREGLPAISLAWGVWE
                     QSGAMTDGLVEADRARMARSGVLPLPPEEGLRLFDAALASDEAVLAPVRIDTGALRAG
                     EAPPVLRALVPAAAARRTARAAAAPSASSLAERLEEMPEAEREKTVLDLVRAEVAAGL
                     GHASDRTVRPEHAFQDLGFDSLTAVELRNRLNKATGLRLPATLVFDHPTPALLARHVF
                     MAAAGAAEPALVDSLLADLDRVEQELVTTLAASEARDRILSKLQGLLAIAGEFGKAPA
                     QEGPGTGGSDLESATDDEVFDLLGKEFGIS"
     gene            21273. .31847
                     /gene="pteA2"
     CDS             21273. .31847
                     /gene="pteA2"
                     /codon_start=1
                     /transl_table=11
                     /product="modular polyketide synthase"
                     /protein_id="BAB69304.1"
                     /db_xref="GI:15824140"
                     /translation="MSCRYPGGVDSPEDLWELVAGGRDAVSEFPADRGWDLDALYDPD
                     PESRGTSYAREGAFLRDLARFDAGLFGISPREALAMDPQQRLLLETSWELFERAGIDA
                     ASLAGSRTGVFAGVMNHEFLAALQNSPQDLEGYLGTGTSGSVASGRVAYTFGLEGPAV
                     TVDTACSSSLVALHLAVQSLRNGECSLAVAGGVTAMVGPATFVGFSRQRGLAPDGRCK
                     PFAAGADGTGWGEGAGLLLVERLSDARRNGHEVLAVVRGSAVNQDGASNGLTAPNGPS
                     QQRVIRQALAAAQLTTAQVDAVEAHGTGTRLGDPIEAQALLATYGQGRPDDKPLWLGS
                     IKSNIGHTQAAAGVAGIIKMVMAMRHGELPRTLHVDAPSVHVDWTAGAVRLLTERTPW
                     SASDQPRRAGISSFGISGTNAHTIIEEAPSAEAFAQDTGAPEGPEEGAESAEPVASEA
                     ARPAAGGPVPWVISGRTEEALRAQAERLLSRTDAEVPPLDVAFSLATTRTALEHRAVL
                     QGETPAELAEGLRAVASGSPARGVARGLSGTGGRVGFLFSGQGSQRLGMGRELYASYP
                     AFAAAYDEVCAQLDAVVDVDSEELHRTGAAQPALFAVEVALFRLLESWGIRADYVAGH
                     SVGELAAAHVAGVLSLDDAAKLVSARARLMQALPAGGAMVAVQATEDEVLPHLTDAVG
                     IAAINGPRSVVVSGAEDAVTAIAEAFRQQGRKSSRLKVSHAFHSPLMDPILEEFAEVV
                     RSLTFHRPQIPVVSNLTGRLAEPYTPQYWVRHVREAVRFADGVRTLNDLGVTTFVEIG
                     PGGVLSALAEGCLDDDVVTVPALRADRPERVALVAAVSELHAHGVSPDWHAFFPGARR
                     VDLPTYAFQYERYWLDASPAAPGGVRAAGLGSADHPLLGAAVSLAGGDERLLTGRLSL
                     RTHPWLSDHAVLGTALLPGTAFVELAVRAADEAGCALLEDLTLEAPLVVPEQGGVAVQ
                     VWVGAVDGTGRRPLTVHSRPEDDSDLPWVRHATGFVTEGSEEAGAAAADRTLTAWPPT

  Query Match           56.7%;  Score 738.6;  DB 1;  Length 78210;
  Best Local Similarity  74.7%;  Pred. No. 1.5e-84;
  Matches  975;  Conservative    0;  Mismatches  309;  Indels   21;  Gaps    3;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || || ||||| ||| | |||||||| ||||| | ||| || || || | ||| ||  |
Db      15790 GAGCCCATCGCGATCGTGGGCATGGCCTGCCGCTACCCCGGTGGCGTGCAGTCACCCGAA 15849

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              ||||||||| ||||||    ||| ||| | |||||| |    |   ||||||   | ||||
Db      15850 GACCTGTGGCAGCTGGTGTTCTCCGGCCGCGACGCCGTGTCGGGCTTCCCCGAGGGCCGC 15909

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||  |  ||| ||||   ||| || | | |||   |||   |||||   |||  |
Db      15910 GGCTGGGACGTGGAGCAGCTCTACGACCCGAATCCCGACCAGTGGGGCACCAGCTACACC 15969

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              |  |  ||||| ||||| |||||||||||| ||||||||||||  |||||||| |||
Db      15970 CGAGAGGGCGGCTTCCTGCACGACGCGGCGGACTTCGACGCCGAGTTCTTCGGGATCTCG 16029

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              || |||||||| |||||||||| |||||||||||| |||||||||||||||| ||||||
Db      16030 CCGCGTGAGGCGCTGGCGATGGACCCGCAGCAGCGGTTGCTGCTGGAGACGTCGTGGGAG 16089

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              |||||||||||||| || ||||| |||  | |||| || ||   | | || || ||||||
Db      16090 GCGTTCGAGCGGGCCGGGATCGACCCGGCGACGGTGCGGGGCAGCAGGACCGGTGTCTTC 16149

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              || ||||   | | |  |||||||||| |||  || ||||| || |  | | |
Db      16150 GCCGGCGTGATGTACCACGACTACGGCGGGCGGGTGCACACCTCGCCCCCAGGGC----- 16204

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
                  |||||||| || || | | ||| |   |||||||| ||   || |||||  ||
Db      16205 -------TGGAGGGCTACCTCGTAAACGGCAGTGCGGGCAGCATCGCCTCCGGCCGGGTC 16257

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
```

```
                  | ||||  |||||||| || |||||||| ||||||| ||||||| ||||||| | |||
Db      16258 TCGTACACGTTCGGGCTCGAGGGGCCGGCGGTCACCGTCGACACGGCGTGCTCGTCCTCG 16317

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              |||||  | || || || ||  ||| ||||||||   || |||||| || | || ||
Db      16318 CTCGTAGCCCTCCACCTCGCTGCGCAGGCGCTGCGGAGCGGCGAGTGCTCGATGGCCCTC 16377

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              | |||||||||| |||| ||| |   ||    | ||||| | |||||| ||| ||||||
Db      16378 GTCGGCGGCGTCACGGTGATGGCAGGCCCGTCCGTGTTCGTGGAGTTCAGCCGTCAGCGC 16437

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              || ||||||| ||||||| |||||||||||| |   || |  || ||||| || ||||||
Db      16438 GGCCTGTCGGCGGACGGCCGGTGCAAGGCGTTCGGCGCCGGGGCGGACGGTACGGGCTGG 16497

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              | |||||||||| || |||||||||||||||||||||||| || ||    |||   |||||
Db      16498 GCCGAGGGCGTGGGCATGCTGTTGGTGGAGCGGTTGTCCGACGCTCGTCGGAACGGGCAC 16557

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              |||||||||||||| ||||| |  ||||| |||||||||||||||||||| ||  ||
Db      16558 AAAGTGCTGGCGGTGGTCCGCGGTACGGCGGTGAACCAGGACGGTGCGTCGAACGGTTTG 16617

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||| |||||||| ||  | ||| ||||||||||||||||||| ||||||||||||
Db      16618 ACGGCTCCGAACGGGCCCTCGCAGCAGCGGGTGATCCGGCAGGCGCTGGCGAACGCGGGC 16677

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||||   ||  || || |||| |||||| ||||| |||| |||| ||||||||||
Db      16678 CTGTCTTCGGATGAGGTCGATGCGGTGGAGGCGCACGGTACGGGTACGAGGCTGGGTGAT 16737

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGC---AGCGGGCCGGTGACAGGCCG 1017
              |||||||||| |||||  || ||||||||||||||    |||| | |  ||| |||||
Db      16738 CCGATCGAGGCGCAGGCGCTGATCGCCACGTACGGGCAGGAGCGGTCGGAGGACCGGCCG 16797

Qy       1018 CTGTGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGT 1077
              |||||||||| || |||||||| |||||||||||  |   ||| || |||||||| |||
Db      16798 TTGTGGCTGGGTTCGCTGAAGTCGAACATCGGGCATTCGCAGGCGGCTGCGGGTGTCGGT 16857

Qy       1078 GGGGTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTG 1137
              || || |||||||||||| |||||| ||| | | || ||| |||||| |||  ||||||||
Db      16858 GGTGTGATCAAGATGGTCATGGCGATGCAGCACGGCGTGCTGCCGCAGACCCTGCATGTG 16917

Qy       1138 GATAAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCG 1197
              ||  ||||| | ||||| |||||||||| ||||| ||||||| |||||||| |||||
Db      16918 GACGAGCCGACCCCGCACGTGGACTGGTCGGCGGGCGCGGTGCAGCTGCTGACCGAGGCC 16977

Qy       1198 GTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGG 1257
              ||| |||||| | |  ||  |    ||  ||| ||||| |  ||||| || |||||
Db      16978 GTGGAGTGGCCTGCGAGCGACG------GGCCGCGCCGGGCCGCGGTGTCCTCCTTCGGT 17031

Qy       1258 ATCGGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
                |  |||||| || ||||| || ||  | ||| ||||  || ||
Db      17032 GTGAGCGGCACCAACGCGCACGTCATCGTCGAGCAGGCCGCGACG 17076
```

```
RESULT 8
AY118081
LOCUS       AY118081               41782 bp    DNA     linear   BCT 27-JUL-2002
DEFINITION  Streptomyces sp. GERI155 putative polyketide type 1 synthase genes
            and unknown gene.
ACCESSION   AY118081
VERSION     AY118081.1  GI:21999182
KEYWORDS    .
SOURCE      Streptomyces sp. GERI155.
  ORGANISM  Streptomyces sp. GERI155
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 41782)
  AUTHORS   Lim,S.K., Nam,D.H., Sohng,J.K. and Yoo,J.C.
  TITLE     Cloning of polyketide type 1 synthase gene of Streptomyces sp.
            GERI155 producing a macrolide antibiotic, GERI155
  JOURNAL   Unpublished
```

```
REFERENCE   2  (bases 1 to 78210)
  AUTHORS   Ikeda,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .78210
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
                     /note="Actinobacteridae"
     CDS             complement(1. .1254)
                     /codon_start=1
                     /transl_table=11
                     /product="putative lipoprotein"
                     /protein_id="BAB69301.1"
                     /db_xref="GI:15824137"
                     /translation="MGEALITLCTRSTRLAATAGVLGGVLAVTALGGTSNAATSGNGA
                     GNHPTSQARINQAGAQEASDARIEITPGQGTYSVGTNHPVSVTVTSGKLTKVTMTAVA
                     TGAEIPGTLSTDGTSWKPNGRLERATKYQIAAEAGDVKGRSTTGSATITTVSPANDFI
                     GHLSPEDGSTVGVAMPVTVNFDKAISDKAAVQSEIRVSSSSGQRVVGHWFNDHRVDFR
                     PENYWRPGSTVTVTLNRHGIQKTVTFKIGRSQISTVDAKTKQMTVVRDGKTIKTIPIS
                     SGSPEHPTYNGQMVISEKFRQIHMNGASVGLTEKNGKPSYDIKAVPHAMRLTDSGTFI
                     HGNYWGTDSVFGEANTSHGCVGLKDVRGDSHGRQPAAWFFDHSIVGDVVIVKNSEEKT
                     MLAPDNGLSDWNMSWSEWVAGSPAT"
     CDS             complement(1256. .2290)
                     /codon_start=1
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="BAB69302.1"
                     /db_xref="GI:15824138"
                     /translation="MPVEFLSHDQVSRYGRFASAPTPGELEQFFRLDAKALESARSKR
                     APANRLGWAVQWGCVRMLGVFPTEDLSVVPEVVVRFAAEQLGVDPGEFAAYGARRQNR
                     YEHAWEIRDAYAYRESPAAEAQVREFLATRVWASWEGPRALFDRAVGWLVDNRVLLPG
                     ITTLTRLVAEVRAQENAALYRTLDEAVPEKLRQSMRDLLKVPEGRRVSELERLRRSPM
                     GVSGSADRLAFSERIRMEAVAMFTAGPGSTDIAKELRVSVQSVQRWRSGLAGDRPGRG
                     AVPQPVSVELSETLFAALDEERAKGFAPSWTIRYLHHDFPIFGQRYAYKWSQGCLHSF
                     GRSAQGYRDQ"
     gene            2935. .21081
                     /gene="pteA1"
     CDS             2935. .21081
                     /gene="pteA1"
                     /codon_start=1
                     /transl_table=11
                     /product="modular polyketide synthase"
                     /protein_id="BAB69303.1"
                     /db_xref="GI:15824139"
                     /translation="MVESELNATGNRDCSASADSADGIAVIGLACRLPGAADPAAFWR
                     LLSEGADAITDVPPDRWDGAVVADASAPDRTGIRRGGFLDSIGDFDAGFFGLSPKEAA
                     AMDPQQRLVLELAWEALEDAHVRAATLRSTRTGVFVGAMWDDYATLHHRSGLTAISPH
                     TVTGLHRSIIANRVSYFLGLNGPSLTVDSGQSSSLVSVHLACESLRKGESTIALAGGV
                     NLNIVPESALGAAKFGGLSPDGRCFTFDARANGYVRGEGGGIVVLKPLARAVADGDPV
                     YCVIRGSAVNNDGGGDGLTVPLQSGQEQVLRLAYERAGVNPADVGYVELHGTGTKVGD
                     PIEAAALGAVLGTGREPGRPLRVGSAKTNVGHLEGAAGITGLLKAVLSLNHRELPASL
                     NFGTPNPAIPLDRLNLRVQTEHSEWAAEDGRPLLAGVSSFGMGGTNCHIVLASGAAVE
                     AGEGATAGGGAAPTFWPLSAKTASGLRAQAAALLAHTEAPLGLALPDVGWSLATGRTA
                     FEHRAVVVGEVRDDFLRALRELSTGGIDAALTTGRAGPRGELAFLFSGQGSQRAGMGR
                     ELAESVPAFASALDEVCAHLDPLLPRPLREVMFAPVGTGEAAELDQTLYTQTSLFAVE
                     VALFRLLESWDITPDVLMGHSIGELAAAHVAGVLSLADACALVAARGRLMQALPAGGA
                     MIAVQAAEDEVRARIGERTDRVSIAALNGPDSVVVSGDEDLATEIADAFAALGRKTSR
                     LRVSHAFHSPHMEPMLAEFRNVAEGLTFHAPRIPIVSNVTGRLSEEPSGYEGCTAAYW
                     VRHVREAVRFADGVARLDEQGVRTYLELGPGGVLASMARAGADSESLFVPALRTRRPE
                     AQALLAAVASLHVHGVEPDWDALFGGGRGARRKVALPTYAFERQRYWLDGDTTSPVAT
                     TGWAPVAPQPDAIAETDSAPVGALGRRLAGLAESERDEAVLDMVRAHIAAVLGHAEAR
                     QVETEWTFKDLGFDSLSSVELRNQLSEATGLRLPSGLLFDHPTPAALARHLAAQSLGG
                     GDAAAAAPATVADPHEPIAIVAMSCRLPGGVSSPEDLWRLLASGGDGISGFPTDRGWD
                     VETLYDPEPGTPGKSSTRHGGFLHEAAEFDAGFFGISPREAAAIDPQQRLVLETAWEA
                     FERAGIDPATLRGSRAGVFIGATAQDYGPRLHEPADGVEGYLLTGTTTSVASGRVAYS
                     FGLEGPAVTVDTACSSSLVALHLAVQSLRSGECTMALAGGVTVMSTPGMFVEFSRQRG
                     LSADGRCKAFAAGADGTGWAEGVGMLLVERLSDARRNGHQVLAVVRGSAVNQDGASNG
                     LTAPNGPSQQRVVRQALAAAGLAPADVDAVEAHGTGTLLGDPIEAEALIAAYGQGRDG
                     DRPLWLGSLKSNIGHTQAAAGVAGVIKMVMAMRHGVLPRTLHVDEPTPHVDWSASGVR
                     LLTEAVEWPARDLPRRAAVSSFGISGTNAHVIVEQAATSEEPARGTAVAGPVPWVVSA
                     KSDVALREQAGRLLSFLDAEGVGGAGVRPEDVALSLATSRAVLERRAAVVGGDLAELR
                     QGLEALASGASGAAPVGGAVGGRVGFLFSGQGSQRIGMGRELYAAYPVFAAAYDEVCA
                     FLDAPVDVDSEELNQTGTTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVA
                     GVLSLEDAAKLVSARAALMQALPDGGAMVAVQATEDEVLRHLTDGVGIAAVNGPRSVV
                     VSGAEDAVTAVGEIFAGRGRKTSRLKVSHAFHSPLMDPMLEEFAAVVRGLTFSESRIP
                     VVSNLTGRLAEVYTPEYWVRHVREAVRFADGIETLGELGVTTFVEVGPGGVLSALAQG
                     CFDDGTDIVTVPVLRADRPEPYALVTAIGRLHTHGISPDWQVLFPGARRVDLPTYAFQ
                     RERYWLDAPRAATAGAPDAMDAEFWDSVESEDRASLGALLGLEPTELDVVASKLSAWR
                     RKRREQSTTDGWRYRITWQPLDDMVRAAPSGTWLYVVSEETAWTEAIRAGLTELGVAL
                     VPFAVGEDTDRDALARAFADAGQEQADGILFAAAQGEAGANALQRLVLLVQALGDAGI
                     EAPLWCLTSGAVSTGTADLLTDPVSAQLWGLGRIVALEQPQRWGGLVDLPAELDSRAL
                     ERLAGLLAQADEDQLAVRASGVYGRRLVRAPLSAAPADAPWSPRGTVLVTGGTGALGG
                     QVARWLAGAGAEHLVLTSRRGPDAPGAAELKAELEELGAQVTVVACDAADRDALAELF
                     GRYIVNAVVHTAGVLDDGLVESLTPERLNHVLRPKVDAALHLHELTRDRQDLDAFVLF
                     SSMTGVWGNGGQGAYGAANAFLDALAEHRRAQGVPALAVAWGSWADGGMADGAAGDHL
                     RRRGVRAIAALPAISVLHGALTHGETSMTVADVDWERFVPAPVGTRPCPLLLGVPEAR
                     EALEARARAAQSPEVPVSALVQRLLGATPGEQGRILLDLVREQAATVLGHTGKGAFEA
                     DRTFRETGFDSLTAVELRHRLNTATGLRLPTTLVFDHPTPTALARHLGEELLGRHGLE
                     TAAAQATAAAAADEPLAIVGMSCRLPGGVQSPEDLWELLSSGRDAMSGFPTDRGWDID
                     ALYDPDPDRPGKTYSREGGFIQGADRFDAGLFGISPREALAMDPQQRLLLETAWEAFE
                     RAGIALASVRASRTGVFIGTNGQDYANGLRNAPEEIEGYALTGKAASVVSGRISYTFG
                     LEGPAVTVDTACSSSLVALHLAAQALRSGECSMALVGGVTVMTTPDLFVEFSRQRGLS
                     PDGRCKAFAAGADGTGWGEGVGLLLVERLSDARRNGHQVLALVRGSAVNQDGASNGLT
                     APNGPSQQRVIRQALASAGLTPADVDAVEAHGTGTRLGDPIEAQALLATYGQEHSDDR
                     PVWLGSVKSNIGHTQAAAGVAGVIKMVLAMRHGVLPQTLHVDEPTAHVDWSAGAVRLL
                     TEPVQWPGTTGPRRAAVSSFGIGGTNAHTIIEEAPAAAAAEPAREHRPVPVPWVLSAK
                     SEVALRAQAERLLAFATDDVSPVDAGFSSATTRSALEHRAAVIGSDPAELRAGLEALA
                     AGEPAANVVAGRAHSADKVGFLFSGQGSQRLGMGRELYAAYPVFAAAYDEVCALLDAT
                     VDVDAETLHQTGSTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVAGVLSL
                     EDAAKLVSARAALMKALPAGGVMVAVQAAEDEMLPHLIDRAGIAAINGPRSVVISGAE
                     DAVTAIAEVFAAQGRKTSRLKVSHAFHSPLMDPMLEEFAEVVRGLTFSEPWIPVVSNL
                     TGRLTESYTPEYWVRHVREAVRFADGVLTLHELGVTTFVEIGPGGVLSALVQGCLDDG
                     TDIVTIPALRADRPEPQALTTAYAQLHVSGAEVDWHAFFPGARRVELPTYAFQRERYW
                     LDAPASAGDMRAVGQGEAGHPLLGAAVPLADSVGHLLTGRLSKHTHPWLVDHAVSGVA
                     LLPGTAFVELAITAADAVGCDLLEELTLETPLLVPERGGVALQVHVGADDGSGRRPLT
                     VHSQIDDGDAYAHEDDSERAWVRHASGVLTVATGGPVDGSLTAWPPVGAEPVDVDGFY
                     DRLAGMALDYGPVFQGLRAAWRSGDDFFAEVELPGAERTDAGAYGLHPALFDAALHTV
                     WLGAVEPEAGTGNGLLPFAWSGVRLAAAGASDLRVKVSRAGTGTSTVSLVLADGTGGP
                     VAEVDSLTLRPVPADQLRAGSGSDDAVFGLKWTPVSLPSAPDGMRIETYEDLAALRAT
                     DPPADTEADAAAAADAVVVPCPTGTGADLATRVREVTHAVLELVQWWLAEERPGRLVL
                     VTRGGDLAQSAVCGLVRSAQTENPDRIVLVETSSDAHAGAGSGVDDADLARVLPGLIA
                     SGEPQAAVHADDEVRVPRLVRVAVPDAEPELPGLGSGTVLLTGASGGLGGLFARHLVA
                     EHGVRSLLLVSRRGGDAPGAAELTADLTAQGADVTWAACDVADRAAVRALLAGPGQSL
                     SAIIHTAGVLDDGIIGSLTPERLDAVFRPKVDAALNLHDLAGELSGDLSAFVLFSSVA
                     GTLGTPGQANYAAANTFLDALAEHRRAAGQRATSLAWGLWAQDSESAMTGSLDHTDLT
                     RIQRMGLAAIPSSDGLRLFDAALALDRAVVAPLRLDLGGLQGTVPPVLRGLRGTVRGT
                     ARRTAQTAPKGSLGQRLAGLAEAEREQVVLDLVRTEVAAVLGHAGAQSVGADDSFKDI
                     GFDSLTAVELRNRLNSAVGMRLPATLIFDYPNPLTLAHFLTGEAAGGEAATAVPTTLS
                     TATGTADEPIAIVGMACRYPGGVQSPEDLWQLVFSGRDAVSGFPEGRGWDVEQLYDPN
                     PDQWGTSYTREGGFLHDAADFDAEFFGISPREALAMDPQQRLLLETSWEAFERAGIDP
                     ATVRGSRTGVFAGVMYHDYGGRVHTSPPGLEGYLVNGSAGSIASGRVSYTFGLEGPAV
                     TVDTACSSSLVALHLAAQALRSGECSMALVGGVTVMAGPSVFVEFSRQRGLSADGRCK
                     AFGAGADGTGWAEGVGMLLVERLSDARRNGHKVLAVVRGTAVNQDGASNGLTAPNGPS
                     QQRVIRQALANAGLSSDEVDAVEAHGTGTRLGDPIEAQALIATYGQERSEDRPLWLGS
                     LKSNIGHSQAAAGVGGVIKMVMAMQHGVLPQTLHVDEPTPHVDWSAGAVQLLTEAVEW
                     PASDGPRRAAVSSFGVSGTNAHVIVEQAATSEVTARGTAVAGPVPWVLSAKSDAALRE
                     QAGRLLSFLDAEGVGGAGVRPEDVAYSLATSRAVLECRAAVVGEELGEFRQGLEGIVS
                     GAAPVGGAVGGRVGFLFSGQGSQRIGMGRELYAAYPVFAAAYDEVCALLDATVDVDAE
                     TLHQTGSTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVAGVLSLEDAAKL
                     VSARAALMQALPAGGAMVAIQATEAEVLPHLTHGVGLAAINGPRSVVLSGSEDAVLDI
                     AEVFAQQGRKTSRLKVSHAFHSPLMDPMLEEFAEVVRGLTFSEPRIPVVSNLTGRLAE
                     VYIPEYWVRHVREAVRFADGLHTLHELGVTMFVEIGPGGVLSALVQGCLDDGTHIVTV
                     PALRADRPEPYALVTAVGQLHTHGISPEWQALFPGARRVDLPTYAFQRERYWLDAPQA
                     SGDVRAAGLGAADHPLLGAAIATADSDGAVLTGLLSLGTHPWLADHAVNGIVLLPGTA
                     FVELAIRAGDETGCGWIEDLTLELPLVVPERGGVTLQVAVGPEDESGRRQLSVHSRVD
                     DRSWLRHATGVLSSAEAPSAAKLGAWPPAGAEAVDLTAFYEGMAAAGLDYGPVFQGLR
                     SVWRDGDDVFAEVELPDESAAEAAAFALHPALLDAALHALAAGGLVSLDDSPALPFAW
                     SGVVVQAAGAAMVRVKLSRTATGSVTLAVADGAGDPVASVESLSLRAASAEQLPGAVG
                     GDQLFGLEWTRFTLPSATDALMIETVADFDALRESEGTPDVVVVPVVPCPAGSGRETA
                     DRVHSTTSEVLALVQWWLAEERAGRLALVTRPGDLAHAAVWGLVRSAQSENPDRIVLV
                     EAEETDEAVRVLPAALASGEPQFAVRDGEVFVPRLVKATRSATGTATGTPDFGVGPVL
```

```
FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI
GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAAL
LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
TLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
KATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD
IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL
```

Query Match 72.6%; Score 944.8; DB 1; Length 64957;
Best Local Similarity 83.6%; Pred. No. 1.9e-110;
Matches 1088; Conservative 0; Mismatches 202; Indels 12; Gaps 1;

```
Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || ||||| || ||  | || |||||  || ||||| ||||| || ||  | || ||| ||
Db      49923 GAGCCGATTGCGATTGTGGGGATGGCGTGTCGTTTTCCGGGTGGGGTGTGTTCGCCGGAG 49864

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              ||  ||||||||||||   || ||||| || || || || ||    || ||  |||  ||
Db      49863 GAGTTGTGGGAGCTGGTTGCGTCGGGTGGGGATGCGATTGGTGAATTTCCGGCCGGTCGG 49804

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||     |||  |      |     ||||||     | | | || ||||  ||  ||
Db      49803 GGGTGGGATCTGGAGGGGTTGTTTGATTCGGACCCTGACCGGTCGGGGACGTCGTACGCG 49744

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              | |   ||||||||  |  | || |||| | | ||||| || | |||||||||| ||||||
Db      49743 CGGTATGGCGGGTTTTTGTATGAGGCGGGGGAGTTCGATGCGGACTTCTTCGGGATCAGT 49684

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              || ||||||||  |||||||||||||||||||||||| || |||||||||||| ||||||
Db      49683 CCGCGTGAGGCGTTGGCGATGGATCCGCAGCAGCGGTTGTTGCTGGAGACGTCGTGGGAG 49624

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              |||||||||||||||| |||||||||||||| | || || |||||||||| ||||||
Db      49623 GCGTTCGAGCGGGCGGGTATCGATCCGCTGTCGATGCGTGGCTCCCGTACGGGTGTCTTC 49564

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              || ||  |  | | |  |||||||||  | ||  ||| |||  .              | ||
Db      49563 GCCGGGGTGATGTACCACGACTACGGATCCCGCCTGGGTACCAT----------CCCC 49516

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              | ||   | |||||| |||||  || |||||  |   |||  ||||  |||||||||  ||
Db      49515 GAGGGATTCGAGGGCTACATCGGCAACGGTAGCGGCGGCGCCGTCGCGTCGGGCCGCGTC 49456

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              |||||||   |||| || || || || ||  || | ||||||||||  || ||| | |||
Db      49455 GCCTACACGCTCGGTCTCGAGGGCCCTGCCGTCTCGGTGGACACGGCATGTTCGTCGTCG 49396

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
```

```
Db      49395 TTGGTGGCGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 49336

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              |||||||| ||  |||| ||||| ||||     | | ||| |||||||||| || ||||||
Db      49335 GCCGGCGGTGTGACGGTGATGTCGACCCCGCACCTCTTCGTCGAGTTCTCACGCCAGCGC 49276

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              || |||||||||||||| | |||||  | | | ||| || ||||||||||||||  |
Db      49275 GGACTGTCGGTGGACGGCCGCTGCAAGTCCTTCGCGGGTGGAGCCGACGGCACCGGCATG 49216

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      49215 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 49156

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              |||||||||||||| | |||||||||||||| |||||||||||||||||||||||| ||
Db      49155 CGGGTGCTGGCGGTGCTGCGCGGCAGTGCGGTCAATCAGGACGGTGCGTCGAATGGGTTG 49096

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      49095 ACGGCGCCGAATGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 49036

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db      49035 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCATGGGACGGGCACGACGCTGGGTGAT 48976

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              |||||||||| ||||||||||||||||||||||||||||||||||||||| ||||||||||
Db      48975 CCGATCGAGGCGCAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTAACAGGCCGCTG 48916

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              |||||||| ||  ||||||| |||||||| ||  |   ||||||||||||||||||||||
Db      48915 TGGCTGGGATCGGTGAAGTCGAACATCGGCCATGCGCAGGCTGCCGCGGGTGTGGGTGGG 48856

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48855 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 48796

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48795 GAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 48736

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db      48735 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGGTC 48676

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              | ||||||||||||||||||||||||||||||||||||||||
Db      48675 AGTGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 48634
```

RESULT 7
AB070949
LOCUS AB070949 78210 bp DNA linear BCT 11-OCT-2001
DEFINITION Streptomyces avermitilis polyene macrolide biosynthetic gene cluster.
ACCESSION AB070949
VERSION AB070949.1 GI:15824136
KEYWORDS .
SOURCE Streptomyces avermitilis DNA.
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
TITLE Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
MEDLINE 21477403

Shirokane, Minato-ku, Tokyo 108-8641, Japan
(E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
Fax:+81-3-3444-6197)

FEATURES Location/Qualifiers

source 1. .64957
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"

gene 101. .12019
/gene="aveA1"

CDS 101. .12019
/gene="aveA1"
/note="contains a loading module, Ave module 1 and module 2
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 1"
/protein_id="BAA84474.1"
/db_xref="GI:5902891"
/translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTR
MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
GDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAG
LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
THHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGAL
ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
TAFVLYSSAAATFGAPGQANYAAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSA
ARELREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
MTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
NLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY
VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGASGG
VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
LRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPE
DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
VSGREVLPWLSGGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
SLMLRWNAPQHPTAESADDDEKPTSATEAEIFKFIDNDLGLS"

gene 12071. .30790
/gene="aveA2"

CDS 12071. .30790
/gene="aveA2"
/note="contains Ave module 3, module 4, module 5 and module 6
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 2"
/protein_id="BAA84475.1"
/db_xref="GI:5902892"
/translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGINP
EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
ACQALRSGECTMALAGGASVMATPFVFTEFSRQRGLAADGRCKAFSAAADGTGWSEGV
GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
NDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALHRL
LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
TQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
NRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLTAVPADEPIAIV
GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
LTYHDYAARFPTAPAGPEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
HLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSAAADGTGWGE
GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAD
LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAADDVPGEAAGDDAGTGGEATGPA
AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH
GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
QGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
PQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
QHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG
TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
GTWQGNGLATGQVSEHLRRRGMPAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA
AVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ

```
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 31422)
  AUTHORS   Omura,S. and Ikeda,H.
  TITLE     Avermectin aglycon synthase gene
  JOURNAL   Patent: JP 2000245457-A 2 12-SEP-2000;
            THE KITASATO INSTITUTE
COMMENT     OS   Streptomyces avermitilis
            PN   JP 2000245457-A/2
            PD   12-SEP-2000
            PF   24-FEB-1999 JP 1999046961
            PR
            PI   SATOSHI OMURA,HARUO IKEDA
            PC   C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC   (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(14643)
            FT   CDS             (14824). .(31419).
FEATURES              Location/Qualifiers
     source           1. .31422
                      /organism="Streptomyces avermitilis"
                      /db_xref="taxon:33903"
BASE COUNT     4136 a  10237 c  11677 g   5372 t
ORIGIN
```

Query Match 72.6%; Score 944.8; DB 6; Length 31422;
Best Local Similarity 83.6%; Pred. No. 2.3e-110;
Matches 1088; Conservative 0; Mismatches 202; Indels 12; Gaps 1;

```
Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              || ||||| || ||  | || ||||| || ||||| ||||| || ||  | || ||| ||
Db      14935 GAGCCGATTGCGATTGTGGGGATGGCGTGTCGTTTTCCGGGTGGGGTGTGTTCGCCGGAG 14994

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              ||  ||||||||||||  || ||||| || || || || || ||    || ||  ||| ||
Db      14995 GAGTTGTGGGAGCTGGTTGCGTCGGGTGGGGATGCGATTGGTGAATTTCCGGCCGGTCGG 15054

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              || |||     |||   |       |    ||||||    | | | || ||||  ||  ||
Db      15055 GGGTGGGATCTGGAGGGGTTGTTTGATTCGGACCCTGACCGGTCGGGGACGTCGTACGCG 15114

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              | |   ||||||||  |  | || |||| | | ||||| || | |||||||| ||||||
Db      15115 CGGTATGGCGGGTTTTTGTATGAGGCGGGGGAGTTCGATGCGGACTTCTTCGGGATCAGT 15174

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              || ||||||||  |||||||||||||||||||||||||| || ||||||||||||| ||||||
Db      15175 CCGCGTGAGGCGTTGGCGATGGATCCGCAGCAGCGGTTGTTGCTGGAGACGTCGTGGGAG 15234

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              |||||||||||||||||  |||||||||||||||  | || || ||||||||||| ||||||
Db      15235 GCGTTCGAGCGGGCGGGTATCGATCCGCTGTCGATGCGTGGCTCCCGTACGGGTGTCTTC 15294

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              || || |   | | |  |||||||||  | ||  |||  |||           | ||
Db      15295 GCCGGGGTGATGTACCACGACTACGGATCCCGCCTGGGTACCAT-----------CCCC 15342

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              | ||   | ||||||  |||||  || |||||  |   |||  ||||  |||||||||  ||
Db      15343 GAGGGATTCGAGGGCTACATCGGCAACGGTAGCGGCGGCGCCGTCGCGTCGGGCCGCGTC 15402

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              |||||||   |||| || || || || ||  || | ||||||||||  || ||| | |||
Db      15403 GCCTACACGCTCGGTCTCGAGGGCCCTGCCGTCTCGGTGGACACGGCATGTTCGTCGTCG 15462

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              | ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15463 TTGGTGGCGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 15522

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              |||||||| ||  |||| ||||| ||||    | | ||| |||||||||| || ||||||
Db      15523 GCCGGCGGTGTGACGGTGATGTCGACCCCGCACCTCTTCGTCGAGTTCTCACGCCAGCGC 15582

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              || ||||||||||||||| | |||||| | | | ||| || ||||||||||||||||  |
Db      15583 GGACTGTCGGTGGACGGCCGCTGCAAGTCCTTCGCGGGTGGAGCCGACGGCACCGGCATG 15642

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      15643 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 15702

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              |||||||||||||| | |||||||||||||||| |||||||||||||||||||||  ||
Db      15703 CGGGTGCTGGCGGTGCTGCGCGGCAGTGCGGTCAATCAGGACGGTGCGTCGAATGGGTTG 15762

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      15763 ACGGCGCCGAATGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 15822

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db      15823 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCATGGGACGGGCACGACGCTGGGTGAT 15882

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              |||||||||| ||||||||||||||||||||||||||||||||||| ||||||||||
Db      15883 CCGATCGAGGCGCAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTAACAGGCCGCTG 15942

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              |||||||| ||  |||||||| |||||||| ||  |  ||||||||||||||||||||||
Db      15943 TGGCTGGGATCGGTGAAGTCGAACATCGGCCATGCGCAGGCTGCCGCGGGTGTGGGTGGG 16002

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16003 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 16062

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16063 GAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 16122

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16123 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGGTC 16182

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              | ||||||||||||||||||||||||||||||||||||||||
Db      16183 AGTGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 16224
```

```
RESULT 6
AB032367/c
LOCUS       AB032367               64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
            Sciences, Kitasato University, Microbial chemistry; 5-9-1
```

```
                     AKTVALPPQEVEQLIGERGGRLWVAAVNGPRSTAVSGDAEAVVEVLAYCAGTGVRPRI
                     PVDYASHCPHVQPLREELLELLGDISPQPYGVPFFSTVEGTWLDTTTLDAAYWYRNLH
                     QPVRFSHDVQALADQGHRVLLEVSPHPTLVPAIEDTTEDTARRRHCDRQPPPRRERHP
                     LLPQRLRLDPYYRHRQTHHVAPLLDPPRHLPPPLDAPRPAHLPFQHQHYWLESSQPGA
                     GSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVV
                     PALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNIL
                     TNLHHHGITPIPLPLTTHHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPH
                     PHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLAR
                     TTLLEHPTHTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTL
                     TPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLT
                     TQLQQKGIHLTITTCDTSTPRPTHNNSLNTIPPQHPVTTVIHTGGILDDATLTNLTPT
                     QLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAH
                     HRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIADTER
                     PYVVIADIDWSKIEHTSQTSDLVSAARAREPEPAVQRPTPPAELHKTLAHQTSADQRAAL
                     LELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFD
                     HPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDF
                     WDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGIS
                     PREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQS
                     TDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMAL
                     AGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRN
                     GHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGT
                     TLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLL
                     PRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEA
                     PADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPG
                     LDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGT
                     GEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQN
                     DNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITGAHLAG
                     ILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSL
                     VISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHT
                     PLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHH
                     NLPNTPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHAQPTPHPHPLDLPTYP
                     FQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLA
                     DHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAAD
                     GKGRRLVSVYARGGSACGGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAV
                     AVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPAL
                     LNGVVQPLSVLLPGGTGFGKGAGFGKGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGG
                     GREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVP
                     VCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQG
                     WLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDT
                     DTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLTRLESSGDVPAQRS
                     GDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLA
                     GVCGVRDLLLVSRRGPDAPGAEGLRRSWPRGAEVRIVACDVGERREVVRLAGGCSCRV
                     SVFVDFSAVASVTVRLPVASDVRKEAAMAYATVEEFTDYLDPDP"
BASE COUNT     1756 a   4059 c   3528 g   1753 t
ORIGIN

  Query Match          86.7%;  Score 1128.8;  DB 1;  Length 11096;
  Best Local Similarity   94.2%;  Pred. No. 2.2e-133;
  Matches 1226;  Conservative    0;  Mismatches   22;  Indels   54;  Gaps    3;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1603 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 1662

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1663 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 1722

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1723 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 1782

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1783 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 1842

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1843 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 1902

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1903 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 1962

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1963 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 2022

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2023 GCGGACGTTGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTAGC 2082

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2083 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 2142

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2143 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGTTCGCGCTG 2202

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              ||||||||||||||||  |  | || ||||||||||||||||||||||||||||||||||
Db       2203 GCCGGCGGCGTCTCGGATGT---CCACCCCGGCATGTTCATCGAGTTCTCCCGGCAGTGT 2259

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              |||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db       2260 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGGTCTCGGCTGCAGCCGACGGCACCGGCTGG 2319

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              |||||||| |||||  ||
Db       2320 GGCGAGGGCATCGGGCTG---------------------------------------- 2337

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2338 ------CTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 2391

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2392 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 2451

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2452 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 2511

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2512 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 2571

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2572 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 2631

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       2632 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTAGAT 2691

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              |||||||||||||||  |   | ||||||||||||||||||||||||||||||||||||
Db       2692 GAGCCGTCGCCGCAGGGG---CTGGCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 2748

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       2749 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTTTCGTCGTTCGGGATC 2808

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              |||||||||||||||||||||||||||||||||||||||||||
Db       2809 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 2850


RESULT 5
E38021
LOCUS       E38021                31422 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Avermectin aglycon synthase gene.
ACCESSION   E38021
VERSION     E38021.1  GI:18626910
KEYWORDS    JP 2000245457-A/2.
```

```
Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2258 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 2317

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       2318 GACCTGTGGGAGCTGCCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 2377

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2378 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 2437

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2438 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 2497

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2498 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 2557

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2558 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 2617

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2618 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 2677

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2678 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 2737

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2738 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 2797

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2798 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 2857

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2858 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 2917

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2918 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 2977

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2978 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 3037

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3038 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 3097

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3098 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 3157

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3158 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 3217

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3218 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 3277

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3278 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 3337

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3338 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 3397

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3398 GAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 3457

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3458 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 3517

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              |||||||||||||||||||||||||||||||||||||||||||
Db       3518 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 3559
```

RESULT 4
AF275943

```
LOCUS       AF275943               11096 bp    DNA     linear   BCT 02-SEP-2000
DEFINITION  Streptomyces avermitilis avermectin polyketide synthase gene,
            partial cds.
ACCESSION   AF275943
VERSION     AF275943.1  GI:9964075
KEYWORDS    .
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 11096)
  AUTHORS   Hong,Y.-S. and Lee,J.J.
  TITLE     Targeted Gene Disruption of the avermectin O-methyltransferase gene
            and polyketide synthase gene from Streptomyces avermitilis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11096)
  AUTHORS   Hong,Y.-S. and Lee,J.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2000) Anticancer Agent Research Laboratory, Korea
            Research Institute of Bioscience and Biotechnology, P.O. Box 116,
            YuSong-Gu, Taejon 305-600, South Korea
FEATURES             Location/Qualifiers
     source          1. .11096
                     /organism="Streptomyces avermitilis"
                     /strain="ATCC31271"
                     /db_xref="ATCC:31271"
                     /db_xref="taxon:33903"
     CDS             220. .>11096
                     /note="Avr1"
                     /codon_start=1
                     /transl_table=11
                     /product="avermectin polyketide synthase"
                     /protein_id="AAG09812.1"
                     /db_xref="GI:9964076"
                     /translation="MADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAE
                     FAPYVDWSVEQVLRDSPDAPGLDRVDVVQRTLFAVMISLAALWRSQGVEPCAVLGHSL
                     GEIAAAHVSGGLSLADAARVGDAWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDN
                     PARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEER
                     VVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGP
                     KTFVEMSPHPVLTMGLQELAADLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGH
                     GETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHL
                     GFDSVMGVELRNRLSKATGLRLPVTLILDHTTPAAVAARLRTAALGHLDEDTAPVPDS
                     PSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQR
                     HAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERA
                     GIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRSAYS
                     FGLEGPAITVDTGCSASLVTLHLACQSLRSGECTFALAGGVSDVHPGMFIEFSRQCGL
                     SVDGRCKAVSAAADGTGWGEGIGLLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQAL
                     ANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMA
                     AAGVGGVIKMVMALREGVLPRTLHVDEPSPQGLAAGAVRLLTEAVPWPGDAAGRLRRA
                     GVSSFGIGGTNAHVILEEAPAAGGCVAGGRVLEGAPGLAISVAESVAAPVAVSAPVAE
                     SVPVPVPVPVPVPVSARSEAGLRAQAEALRQYEAVQPDVSLADVGAGLACRQAVLEHH
                     VVILAACTSSSRSAARTTARSSSTARPQARPAPTSARRIGALAAGSGSAALTTGHAPG
                     GDRGGVVFVFPWQGGQWAGMGVRLLCLLRVFARRMQACEEALAPWVDWSVVDILRRDA
                     GDAVWEQADVVQPVLFSVMVSLAALWRSYGIEPNEVLGHSKDEIAAAHIYGALSLKDA
```

```
DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAAL
LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
TLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
KATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD
IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL
```

Query Match 100.0%; Score 1302; DB 1; Length 64957;
Best Local Similarity 100.0%; Pred. No. 2.3e-155;
Matches 1302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1541 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 1600

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1601 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 1660

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1661 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 1720

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1721 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 1780

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1781 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 1840

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1841 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 1900

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1901 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 1960

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1961 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 2020

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2021 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 2080

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2081 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 2140

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2141 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 2200

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2201 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 2260

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2261 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 2320

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2321 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 2380

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2381 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 2440

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2441 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 2500

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2501 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 2560

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2561 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 2620

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2621 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 2680

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2681 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 2740

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2741 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 2800

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              |||||||||||||||||||||||||||||||||||||||||||
Db       2801 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 2842
```

```
RESULT 3
AX006889
LOCUS       AX006889               12381 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 1 from Patent WO0001827.
ACCESSION   AX006889
VERSION     AX006889.1  GI:9994904
KEYWORDS    .
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 12381)
  AUTHORS   Kellenberger,J.L., Leadlay,P.F., Staunton,J., Mcarthur,H.A. and
            Stutzman-Engwall,K.J.
  TITLE     Polyketides, their preparation, and materials for use therein
  JOURNAL   Patent: WO 0001827-A 1 13-JAN-2000;
            KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
            STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PFIZER (US); MCARTHUR
            HAMISH ALASTAIR IRVIN (US); STUTZMAN ENGWALL KIM JONELLE (US)
FEATURES             Location/Qualifiers
     source          1. .12381
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     1884 a    4561 c    4005 g    1931 t
ORIGIN
```

Query Match 99.8%; Score 1298.8; DB 6; Length 12381;
Best Local Similarity 99.8%; Pred. No. 9.2e-155;
Matches 1300; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
                /gene="aveA1"
CDS             101. .12019
                /gene="aveA1"
                /note="contains a loading module, Ave module 1 and module
                2
                multifunctional polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="type I polyketide synthase AVES 1"
                /protein_id="BAA84474.1"
                /db_xref="GI:5902891"
                /translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
                ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
                AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
                VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTR
                MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
                MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
                RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
                DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
                LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
                GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
                MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
                GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
                GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
                RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
                GDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
                LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
                AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAG
                LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
                EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
                VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
                AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
                AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
                ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
                PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
                THHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
                EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
                WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
                LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
                PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
                LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGAL
                ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
                QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
                TAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
                DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSA
                ARAREREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
                FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
                AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
                YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
                GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
                GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
                GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
                NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
                WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
                TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
                SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
                ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
                YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
                VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
                MTTLHTTPHHITHHLTAHENDLAIAAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
                LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
                VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
                NLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
                TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY
                VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGG
                VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
                RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
                EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
                RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
                LRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPE
                DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
                ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
                VSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
                AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
                LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
                RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
                RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
                TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
                AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
                SLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLGLS"
gene            12071. .30790
                /gene="aveA2"
CDS             12071. .30790
                /gene="aveA2"
                /note="contains Ave module 3, module 4, module 5 and
                module 6
                multifunctional polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="type I polyketide synthase AVES 2"
                /protein_id="BAA84475.1"
                /db_xref="GI:5902892"
                /translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
                CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
                DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGINP
                EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
                ACQALRSGECTMALAGGASVMATPFVFTEFSRQRGLAADGRCKAFSAAADGTGWSEGV
                GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
                PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
                GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
                GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
                KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
                ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
                NDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALHRL
                LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
                TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
                FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
                TQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
                TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
                PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
                NRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLTAVPADEPIAIV
                GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
                FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
                LTYHDYAARFPTAPAGFEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
                HLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSAAADGTGWGE
                GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAD
                LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
                VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
                SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAADDVPGEAAGDDAGTGGEATGPA
                AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
                IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH
                GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
                PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
                QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
                QGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
                TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
                PQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
                SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
                QHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
                PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
                LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
                IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG
                TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
                TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
                LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
                GTWQGNGLATGQVSEHLRRRGMFAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
                LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA
                AVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ
                FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI
                GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
                RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
                SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
                TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
                AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
                LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
                HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
```

```
            THE KITASATO INSTITUTE
COMMENT     OS   Streptomyces avermitilis
            PN   JP 2000245457-A/1
            PD   12-SEP-2000
            PF   24-FEB-1999 JP 1999046961
            PR
            PI   SATOSHI OMURA,HARUO IKEDA
            PC   C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC   (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(11916)
            FT   CDS             (11971). .(30687).
FEATURES             Location/Qualifiers
     source          1. .30690
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     5355 a  12455 c   8617 g   4263 t
ORIGIN

  Query Match           100.0%;  Score 1302;  DB 6;  Length 30690;
  Best Local Similarity  100.0%;  Pred. No. 2.9e-155;
  Matches 1302;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GACCCGATCGCCATCATCGGCATGGCATGCCGTTTCCCGGGCGGAGTCCGGTCCCCGAAG 1500

Qy         61 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 GACCTGTGGGAGCTGGCCGCCTCGGGCGGAGACGCCATCGGGCCGTTCCCCACCGACCGC 1560

Qy        121 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GGATGGCCCACGGAACAGCGTCACGCCCAGGACCCCACGCAGCCCGGCACGTTCTATCCG 1620

Qy        181 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 CAGGGAGGCGGGTTCCTTCACGACGCGGCGCACTTCGACGCCGGCTTCTTCGGAATCAGT 1680

Qy        241 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 CCACGTGAGGCACTGGCGATGGATCCGCAGCAGCGGCTGCTGCTGGAGACGTCCTGGGAG 1740

Qy        301 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GCGTTCGAGCGGGCGGGAATCGATCCGCTGTCGGTACGCGGGTCCCGTACGGGCGTCTTC 1800

Qy        361 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GCGGGCGCCCTCTCCTTCGACTACGGCCCGCGTATGGACACCGCGTCGTCGGAGGGCGCC 1860

Qy        421 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 GCGGACGTGGAGGGCCACATCCTCACCGGTACCACGGGCAGCGTCCTGTCGGGCCGTATC 1920

Qy        481 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 GCCTACAGCTTCGGGCTGGAAGGGCCGGCGATCACCGTGGACACGGGGTGCTCGGCATCG 1980

Qy        541 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 CTCGTGACGCTGCATCTGGCGTGCCAGTCGCTGCGGTCGGGTGAGTGCACGCTCGCGCTG 2040

Qy        601 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 GCCGGCGGCGTCTCGGTCATGTCCACCCTCGGCATGTTCATCGAGTTCTCCCGGCAGCGC 2100

Qy        661 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GGGCTGTCGGTGGACGGCAGGTGCAAGGCGTACTCGGCTGCAGCCGACGGCACCGGCTGG 2160

Qy        721 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 GGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGGTTGTCGGATGCGGTGCGGCTGGGGCAT 2220

Qy        781 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTCAACCAGGACGGTGCGTCGAATGGGCTG 2280

Qy        841 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 ACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTGATCCGGCAGGCGTTGGCGAACGCGGGG 2340

Qy        901 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 TTGTCCGTGGCGGATGTGGATGTGGTGGAGGGGCACGGGACGGGCACGACGCTGGGTGAT 2400

Qy        961 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 CCGATCGAGGCACAGGCGTTGCTCGCCACGTACGGGCAGCGGGCCGGTGACAGGCCGCTG 2460

Qy       1021 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 TGGCTGGGGTCTCTGAAGTCCAACATCGGGCACACCATGGCTGCCGCGGGTGTGGGTGGG 2520

Qy       1081 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 GTCATCAAGATGGTGATGGCGTTGCGGGAGGGGGTGTTGCCGCGGACGTTGCATGTGGAT 2580

Qy       1141 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 AAGCCGTCGCCGCAGGTGGACTGGTCCGCGGGGGCGGTGCGGCTGCTGACGGAGGCGGTG 2640

Qy       1201 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 CCGTGGCCGGGGGACGCGGCAGGGCGGTTGCGGCGGGCGGGAGTGTCGTCGTTCGGGATC 2700

Qy       1261 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 1302
              |||||||||||||||||||||||||||||||||||||||||||
Db       2701 GGCGGCACGAATGCGCATGTGATTTTGGAGGAGGCGCCGGCG 2742


RESULT 2
AB032367
LOCUS       AB032367               64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
            Sciences, Kitasato University, Microbial chemistry; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .64957
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            101. .12019
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 2525.62 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_1441_2742
Perfect score: 1302
Sequence: 1 gacccgatcgccatcatcgg.....ttttggaggaggcgccggcg 1302

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1302	100.0	30690	6	E38020	E38020 Avermectin
2		1302	100.0	64957	1	AB032367	AB032367 Streptomy
3		1298.8	99.8	12381	6	AX006889	AX006889 Sequence
4		1128.8	86.7	11096	1	AF275943	AF275943 Streptomy
5		944.8	72.6	31422	6	E38021	E38021 Avermectin
6	c	944.8	72.6	64957	1	AB032367	AB032367 Streptomy
7		738.6	56.7	78210	1	AB070949	AB070949 Streptomy
8		722.2	55.5	41782	1	AY118081	AY118081 Streptomy
9		710	54.5	20394	1	SNA132222	AJ132222 Streptomy
10		710	54.5	20394	6	AX067996	AX067996 Sequence
11	c	710	54.5	84985	1	SNA278573	AJ278573 Streptomy
12		689	52.9	24568	1	AF220951	AF220951 Streptomy
13		689	52.9	50937	6	AR159871	AR159871 Sequence
14	c	688	52.8	104326	1	AB070940	AB070940 Streptomy
15		677	52.0	30000	6	AX250262	AX250262 Sequence
16		676.4	52.0	43280	1	SFU78289	U78289 Streptomyce
17		675.8	51.9	104326	1	AB070940	AB070940 Streptomy
18		674.6	51.8	25150	1	AB070945	AB070945 Streptomy
19		672.6	51.7	30000	6	AX250263	AX250263 Sequence
20		662.6	50.9	5544	1	SNA132221	AJ132221 Streptomy
21		662.6	50.9	5544	6	AX067994	AX067994 Sequence
22		662.6	50.9	84985	1	SNA278573	AJ278573 Streptomy
23		659.8	50.7	11480	1	STMPKS3ORF	L09654 Streptomyce
24		655	50.3	53784	1	AMM223012	AJ223012 Amycolato
25		655	50.3	53789	6	A69720	A69720 Sequence 3
26		655	50.3	90445	1	AF040570	AF040570 Amycolato
27		652.8	50.1	3900	1	AF262754	AF262754 Amycolato
28		650.4	50.0	14775	6	AX089462	AX089462 Sequence
29		650.4	50.0	50000	6	AX089417	AX089417 Sequence
30		650.4	50.0	50000	6	AX089420	AX089420 Sequence
31		650.4	50.0	80161	1	AY007564	AY007564 Saccharop
32		650.4	50.0	80161	6	AR165018	AR165018 Sequence
33		649	49.8	16767	6	AX089464	AX089464 Sequence
34		646.4	49.6	9513	6	AX089460	AX089460 Sequence
35	c	645.8	49.6	24225	1	SC2C4	AL512902 Streptomy
36		645.4	49.6	6459	6	AX089458	AX089458 Sequence
37	c	641.4	49.3	82746	1	AF453501	AF453501 Actinosyn
38		639	49.1	37948	1	AF079138	AF079138 Streptomy
39		638.2	49.0	27541	6	AX211706	AX211706 Sequence
40		638.2	49.0	123580	1	AF263912	AF263912 Streptomy
41		638.2	49.0	125401	6	AX211739	AX211739 Sequence
42	c	637.4	49.0	107379	1	SHGCPIR	X86780 S.hygroscop
43		636.4	48.9	41097	1	AF016585	AF016585 Streptomy
44	c	636.2	48.9	26195	1	SC1G7	AL591083 Streptomy
45		636	48.8	65140	6	AX211705	AX211705 Sequence

ALIGNMENTS

RESULT 1
E38020
LOCUS E38020 30690 bp DNA linear PAT 31-JAN-2002
DEFINITION Avermectin aglycon synthase gene.
ACCESSION E38020
VERSION E38020.1 GI:18626909
KEYWORDS JP 2000245457-A/1.
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 30690)
AUTHORS Omura,S. and Ikeda,H.
TITLE Avermectin aglycon synthase gene
JOURNAL Patent: JP 2000245457-A 1 12-SEP-2000;

THIS PAGE BLANK (USPTO)

```
Qy        245 TC 246
                |
Db        448 CC 447
```

Search completed: June 18, 2003, 01:00:34
Job time : 318.438 secs

```
Db        332 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCTCCCCCCCCCCCCCCC 391

Qy        185 TCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCACCACC 244
              || || |||| || | | | ||  |  || ||||  |||| || ||| | ||| || ||
Db        392 CCCCCCCCCCCCCCTCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCC 451

Qy        245 TC 246
              ||
Db        452 CC 453


RESULT 14
CNS0118A
LOCUS       CNS0118A                 848 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN05N21 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL099940
VERSION     AL099940.1  GI:5611551
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 848)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .848
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN05N21"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
BASE COUNT      119 a    518 c     91 g     75 t     45 others
ORIGIN

  Query Match          32.0%;  Score 78.8;  DB 17;  Length 848;
  Best Local Similarity   50.4%;  Pred. No. 6.2e-05;
  Matches  122;  Conservative   24;  Mismatches   96;  Indels    0;  Gaps    0;

Qy          5 CCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCACATCG 64
              |:  ||    | |::| :: ::: |:  || :|| || |::||  | |::| ||| |
Db         62 CMCMCCCCCCCCMMGCMMYMMMYCMGCCCCSCCCCCSCSSCCCCGCCMSCCCCCSCCCGA 121

Qy         65 CCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGCGACC 124
              || ||| ||   ||| | :| ||||| | ||  |||||||  :| | ||  || | |:
Db        122 CCCCCGCCCCACGCCCCCSCCCCCCCCCCGCCCCCCCCCCCCCMCCCCCGCCCCCCCCCM 181

Qy        125 TCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCGGAC 184
              |  |  |  | ||| | |: ||  |  ||    :| ||  | | ||  |||| ||   |
Db        182 CCCCGGCCCCCMCCCCCSCMCCCCCCCGACCCACMAGCCCCCCCCCCGCCACCCCCCCCC 241

Qy        185 TCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCACCACC 244
              ||||| :|||||  |  | ||  |  ||::||| |||| || ||  | ::| || ||
Db        242 CCCGCCCSCCCACCCACGCCCCCCCCCCCCMMCCCCCCCCCCCCCCCCACCCMMCCCCCCC 301

Qy        245 TC 246
              ||
Db        302 CC 303


RESULT 15
AZ687957/c
LOCUS       AZ687957                 913 bp    DNA     linear   GSS 14-DEC-2000
DEFINITION  ENTIJ61TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ687957
VERSION     AZ687957.1  GI:11825103
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 913)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 16
            High quality sequence stop: 292.
FEATURES             Location/Qualifiers
     source          1. .913
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT       90 a     63 c    438 g    322 t
ORIGIN

  Query Match          32.0%;  Score 78.8;  DB 17;  Length 913;
  Best Local Similarity   57.9%;  Pred. No. 6.1e-05;
  Matches  140;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy          5 CCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCACATCG 64
              ||  ||   | |||    |  || ||| || ||| || | |||    || | ||| |  ||
Db        688 CCCCCCCCCCCCACCCCCCCCACCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCG 629

Qy         65 CCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGCGACC 124
              ||  ||   ||    || | || |||||  | ||  |||||||  || | |    || |  ||
Db        628 CCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 569

Qy        125 TCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCGGAC 184
              |  |  |  | ||| | || ||  |   | | |  ||  | | ||| | || ||  ||
Db        568 CCCCCCCCCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCAC 509

Qy        185 TCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCACCACC 244
              || || |||| || ||| | ||  |  || ||||  |||| || ||| | ||| || ||
Db        508 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 449
```

```
Db        263 CCCNCCNCCCCCCCNNCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCNNNCCCCCCCCCCC 204

Qy        245 TC 246
              |
Db        203 CC 202
```

RESULT 12
BG519974
LOCUS BG519974 375 bp mRNA linear EST 10-MAY-2001
DEFINITION ps07c09.y2 Trichinella spiralis ML CMVsport jasmer Trichinella spiralis cDNA 5' similar to contains element MSR1 repetitive element ;, mRNA sequence.
ACCESSION BG519974
VERSION BG519974.1 GI:13535401
KEYWORDS EST.
SOURCE Trichinella spiralis.
ORGANISM Trichinella spiralis
Eukaryota; Metazoa; Nematoda; Enoplea; Trichocephalida; Trichinellidae; Trichinella.
REFERENCE 1 (bases 1 to 375)
AUTHORS McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagareishvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Dr. Doug Jasmer (djasmer@vetmed.wsu.edu) at Washington State University, Dept. of Veterinary Microbiology and Pathology DNA Sequencing by: Washington University Genome Sequencing Center St. Louis.
Seq primer: Sp6
High quality sequence stop: 371.
FEATURES Location/Qualifiers
source 1. .375
/organism="Trichinella spiralis"
/db_xref="taxon:6334"
/clone_lib="Trichinella spiralis ML CMVsport jasmer"
/dev_stage="muscle stage larvae"
/lab_host="DH10B"
/note="Vector: pCMVsport-7.neo; Site_1: NotI; Site_2: SalI ; The library was constructed using mRNA isolated from total RNA with oligo-dT cellulose. Total RNA was generated from muscle larvae that were isolated from infected rats. Larvae were liberated by pepsin/HCl digestion, incubated with 1% SDS, treated with RNase and DNase to eliminate host nucleic acid contamination, and purified on a Percoll gradient. The T. spiralis isolate was obtained from Dickson Despommier (Columbia University). The library was provided by Dr. Doug Jasmer (djasmer@vetmed.wsu.edu) and colleagues at Washington State University. DNA Sequencing by: Washington University Genome Sequencing Center St. Louis."
BASE COUNT 17 a 294 c 10 g 7 t 47 others
ORIGIN

Query Match 32.0%; Score 78.8; DB 12; Length 375;
Best Local Similarity 56.6%; Pred. No. 6.8e-05;
Matches 137; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
Qy          5 CCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCACATCG 64

Db        128 CCCCCCCCCCCAAACCAACACCAAAGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 187

Qy         65 CCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGCGACC 124

Db        188 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 247

Qy        125 TCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCGGAC 184

Db        248 CCCCCNCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 307

Qy        185 TCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCACCACC 244

Db        308 CCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCNCCCCCNCNNNCCCCC 367

Qy        245 TC 246
              |
Db        368 CC 369
```

RESULT 13
AW335294
LOCUS AW335294 585 bp mRNA linear EST 31-JAN-2000
DEFINITION S45E8 AGS-1 Pneumocystis carinii f. sp. carinii cDNA 3', mRNA sequence.
ACCESSION AW335294
VERSION AW335294.1 GI:6831913
KEYWORDS EST.
SOURCE Pneumocystis carinii f. sp. carinii.
ORGANISM Pneumocystis carinii f. sp. carinii
Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes; Pneumocystidaceae; Pneumocystis.
REFERENCE 1 (bases 1 to 585)
AUTHORS Smulian,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C., Edman,J.C., Kovacs,J. and Cushion,M.
TITLE Expressed sequence tags from Pneumocystis carinii
JOURNAL Unpublished (2000)
COMMENT Contact: Staben C
School of Biological Sciences
University of Kentucky
101 Morgan Building, University of Kentucky, Lexington, KY 40506-0225, USA
Tel: 606 257 2161
Fax: 606 257 1717
Email: staben@pop.uky.edu.
FEATURES Location/Qualifiers.
source 1. .585
/organism="Pneumocystis carinii f. sp. carinii"
/db_xref="taxon:38081"
/clone_lib="AGS-1"
/lab_host="E. coli"
/note="Vector: Lambda ZAP II; Site_1: EcoRI; Site_2: XhoI; P. carinii organisms (3x10e9) from a single rat (99-1-6, sacrificed on 3/17/99) at Cincinnati VA facilities. Trizol extracted RNA. Oligo dT priming, standard conditions described by vendor, Stratagene. Further details see www.uky.edu/Project/Pneumocystis/"
BASE COUNT 0 a 565 c 3 g 17 t
ORIGIN

Query Match 32.0%; Score 78.8; DB 10; Length 585;
Best Local Similarity 57.9%; Pred. No. 6.5e-05;
Matches 140; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

```
Qy          5 CCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCACATCG 64

Db        212 CCCCCCTCCCCCCCCCCCCCCCCGCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCC 271

Qy         65 CCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGCGACC 124

Db        272 CCCCCCCCCTGCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 331

Qy        125 TCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCGGAC 184
```

```
Db        425 CCCCCCCACCCCCCCCCCCANCCCCCCCCCCCCCCCCCCCCACCCCCCCACNCCCCCCCCC 366

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              | || || || | || ||| |  |  |  || |||||  ||  || ||| | ||| |
Db        365 CCCCCCCCCCCCCACCCCCCCCCCCNCCCCCCCCCCCCCCCACACCACCCCCCCCCCCCCC 306

Qy        241 CACCTC 246
              | ||| |
Db        305 CCCCCC 300


RESULT 10
AG153752/c
LOCUS       AG153752                 744 bp    DNA     linear   GSS 09-JAN-2002
DEFINITION  Pan troglodytes DNA, clone: RP43-017I24.TJ, genomic survey
            sequence.
ACCESSION   AG153752
VERSION     AG153752.1  GI:16683430
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphocytes DNA, clone_lib:RPCI-43 Chimpanzee
            Male BAC Library clone:RP43-017I24.TJ.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library RPCI-43
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 744)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
            end was generated during the R&D process and may have higher chance
            of clone tracking errors.
             PRIMERS
                 Sequencing: TJ
             LIBRARY
                 Vector    : pBACe3.6
                 R.Site 1  : EcoRI
                 R.Site 2  : EcoRI.
FEATURES             Location/Qualifiers
     source          1. .744
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="RP43-017I24.TJ"
                     /sex="male"
                     /cell_type="lymphocytes"
                     /clone_lib="RPCI-43 Chimpanzee Male BAC Library"
BASE COUNT       27 a     52 c    515 g     13 t    137 others
ORIGIN

  Query Match             32.1%;  Score 79;  DB 17;  Length 744;
  Best Local Similarity   54.3%;  Pred. No. 5.8e-05;
  Matches  133;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy          2 TGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCACA 61
              | || |||| | | | |   |  |  | | || ||| || | |||    || | |||
Db        729 TTGCGGGCCCGCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCNC 670

Qy         62 TCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGCG 121
               | || ||  |     || | || |||||  | ||  |||||||   |   |   || |
Db        669 CCCCCCCCCNCNNCCNCCCCCCCCCCCCCCCCCNCCNNCCCCCCCCCNCNNCCNCCCCCCC 610

Qy        122 ACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCG 181
              |  |  |  |  | ||| | || ||  |   |  | |  ||  | | ||  | || ||
Db        609 CCNNCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCNCNCCCCCCCCCCCNCCNCCCCCC 550

Qy        182 GACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCACC 241
              | || || |||| || ||| | ||  |  || ||||  |||| || ||| | ||| ||
Db        549 CCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 490

Qy        242 ACCTC 246
              || |
Db        489 CCCCC 485


RESULT 11
AG136783/c
LOCUS       AG136783                 912 bp    DNA     linear   GSS 04-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-150H12.F, genomic survey sequence.
ACCESSION   AG136783
VERSION     AG136783.1  GI:16666461
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-150H12.F.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 912)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .912
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-150H12.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      177 a     26 c    496 g    110 t    103 others
ORIGIN

  Query Match             32.1%;  Score 79;  DB 17;  Length 912;
  Best Local Similarity   55.0%;  Pred. No. 5.7e-05;
  Matches  133;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy          5 CCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCACATCG 64
              ||  ||   |     |||| |  | | || ||| || | |||    || | ||| |  |
Db        443 CCNNCCNCNCCACAAAACACCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCC 384

Qy         65 CCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGCGACC 124
              || ||  ||    || | || |||||  | ||  ||||||| || | |   || |  ||
Db        383 CCCCCNCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 324

Qy        125 TCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCGGAC 184
              |  | | || ||| | || ||  |   | | |  ||  | | ||| | || ||   |
Db        323 CCNCCCCCCACCCCCNCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCNCC 264

Qy        185 TCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCACCACC 244
```

```
Qy          5 CCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCACATCG 64
              ||  ||   | | |   || |  | | || ||| || | |||    || | ||| |  |
Db        834 CCCCCCCCCCCCCCCCCACCCCCCCCCNCCCCCCCCCCCCCCCCACCCCCCCCCCCCCC 775

Qy         65 CCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGCGACC 124
              || ||  ||    || | || |||||  | ||  |||||||  || | |   || |  ||
Db        774 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 715

Qy        125 TCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCGGAC 184
              |  | | || ||| | || ||   |   | | |  ||  | | ||| | |||||   |
Db        714 CCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 655

Qy        185 TCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCACCACC 244
              || || ||||||| ||| | ||  |  |||||||  |||| || ||| | ||| || ||
Db        654 CCCCCCCCCCCACCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCC 595

Qy        245 TC 246
              |
Db        594 CC 593
```

RESULT 8
CNS04FIQ
LOCUS CNS04FIQ 389 bp DNA linear GSS 21-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 106L07 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL288395
VERSION AL288395.1 GI:8026928
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 389)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 389)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
TITLE Charaterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 389)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000)
COMMENT This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
source 1. .389
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="106L07"
/clone_lib="G"
/note="Genoscope sequence ID : C0BG106CF04SP1~end : PUC-Ori"
BASE COUNT 34 a 279 c 34 g 29 t 13 others
ORIGIN

Query Match 32.2%; Score 79.2; DB 17; Length 389;
Best Local Similarity 55.0%; Pred. No. 5.8e-05;
Matches 132; Conservative 10; Mismatches 98; Indels 0; Gaps 0;

```
Qy          7 GGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCACATCGCC 66
              ||  | ||||:|  :|:  | |:|:|: :|: :| | ||| |  |||   || |  | ||
Db        117 GGTTACACACMCACMCMCAAACMCMCMCMCMCACACCCGGCCCCGACTCCCCCCCCCC 176

Qy         67 ACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGCGACCTC 126
              ||  ||    || | || |||||  | ||  |||||||  || | |   || |  || |
Db        177 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCYCCCCCCCCCTCCCCC 236

Qy        127 GGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCGGACTC 186
              |  | | ||||| :| ||  |   |  | |  ||  | | ||| | || ||   | |
Db        237 CCCCCCCCCCCCCTCCYCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCC 296

Qy        187 CGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCACCACCTC 246
              | || |||| || ||| |  || |||| |||| || ||| | ||| || || | |
Db        297 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 356
```

RESULT 9
BG786282/c
LOCUS BG786282 646 bp mRNA linear EST 20-MAY-2001
DEFINITION SEAUMC006239 Sea urchin primary mesenchyme cell cDNA library Strongylocentrotus purpuratus cDNA clone PC_0028_A2_C11_MR 5', mRNA sequence.
ACCESSION BG786282
VERSION BG786282.1 GI:14157295
KEYWORDS EST.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.
REFERENCE 1 (bases 1 to 646)
AUTHORS Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and Ettensohn,C.A.
TITLE A large scale analysis of mRNAs expressed by primary mesenchyme cells of the sea urchin embryo
JOURNAL Development 128 (13), 2615-2627 (2001)
MEDLINE 21384984
COMMENT Contact: Ettensohn CA
Dept. Biol. Sci.
Carnegie Mellon University
4400 Fifth Avenue, Pittsburgh, PA 15213, USA
Tel: +1 412 268 5849
Email: ettensohn@andrew.cmu.edu.
FEATURES Location/Qualifiers
source 1. .646
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="PC_0028_A2_C11_MR"
/clone_lib="Sea urchin primary mesenchyme cell cDNA library"
/tissue_type="embryo"
/cell_type="primary mesenchyme cells"
/lab_host="E.coli"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; oligo dt priming from poly A+ RNA, directionally cloned"
BASE COUNT 2 a 3 c 566 g 50 t 25 others
ORIGIN

Query Match 32.1%; Score 79; DB 12; Length 646;
Best Local Similarity 56.5%; Pred. No. 5.9e-05;
Matches 139; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

```
Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              |   ||  ||   | | |   || || |||| | ||| || | | |||    || | ||| |
Db        545 CNCNCCCCCCCCCCCCCCCCCCCACCAACACANCCCCCCCCCCCCCCCCCCCCCCCCCC 486

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
              | |||||  ||     |||| || ||||| || ||| |||||||  || | |   || |
Db        485 CCCCCCACCCCCCCCCCCCACCCCCCCCCCCACCCCACCCCCCCCCCCCCCNCCCACCCC 426

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              || |  |  |  | |||   || ||  |   |  | | |||  | | | | | || ||
```

```
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="22N11"
                     /clone_lib="NotreDame1"
                     /note="end : T7"
BASE COUNT       136 a    552 c     71 g    112 t     83 others
ORIGIN

  Query Match             32.4%;  Score 79.6;  DB 17;  Length 954;
  Best Local Similarity   56.2%;  Pred. No. 4.4e-05;
  Matches  136;  Conservative    6;  Mismatches  100;  Indels    0;  Gaps    0;

Qy          5 CCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCACATCG 64
              ||  |||  ||: | : |  | ||:|:|| ||| |  | |||    ||   ||| |  |:
Db        175 CCCMCCACCCAMCCCMCCCCCCGCMCMCCCCCCCCGMCMCCCMMCMCCCGSCCCCCCCCS 234

Qy         65 CCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGCGACC 124
              || ||  ||  |:|| | || |||||  | ||  |||||||  || | |   ||  | ||
Db        235 CCCCCCCCCCCGSCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGGCCC 294

Qy        125 TCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCGGAC 184
              |  |  |  | ||| | || ||  |   |  | |  ||  | | ||| | || ||   |
Db        295 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACC 354

Qy        185 TCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCACCACC 244
              ||| || |||| || ||| | ||  |  || |||  |||| || ||| | ||| || ||
Db        355 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 414

Qy        245 TC 246
              |
Db        415 CC 416


RESULT 6
BI416474/c
LOCUS       BI416474                803 bp    mRNA    linear   EST 15-AUG-2001
DEFINITION  hasp001xa09f Heterobasidion annosum - Scots pine infection stage
            subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
            annosum cDNA clone hasp001xa09f, mRNA sequence.
ACCESSION   BI416474
VERSION     BI416474.1  GI:15187497
KEYWORDS    EST.
SOURCE      Pinus sylvestris/Heterobasidion annosum.
  ORGANISM  Pinus sylvestris/Heterobasidion annosum
            Eukaryota; mixed EST libraries.
REFERENCE   1  (bases 1 to 803)
  AUTHORS   Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
  TITLE     Expressed sequence tags of randomly selected cDNA clones from the
            interaction of the root rot fungus (Heterobasidion annosum) with
            seedling roots of Scots pine (Pinus sylvestris)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Fred O. Asiegbu
            Dept. of Forest Mycology & Pathology
            Swedish University of Agriculture, Box 7026,S-750 07, Uppsala,
            Sweden
            Tel: +46 18 67 15 98
            Fax: +46 18 30 92 45
            Email: Fred.Asiegbu@mykopat.slu.se
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .803
                     /organism="Pinus sylvestris/Heterobasidion annosum"
                     /db_xref="taxon:169015"
                     /clone="hasp001xa09f"
                     /clone_lib="Heterobasidion annosum - Scots pine infection
                     stage subtraction cDNA library (hasp)"
                     /dev_stage="Seedling roots of scots pine were infected for
                     6 days with H. annosum"
                     /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive
                     hybridization cDNA library was constructed from  scots
                     pine roots infected for 6-days with mycelia of
                     Heterobasidon  annosum (FP5)."
BASE COUNT       45 a     17 c    723 g     13 t      5 others
ORIGIN

  Query Match             32.3%;  Score 79.4;  DB 13;  Length 803;
  Best Local Similarity   57.9%;  Pred. No. 4.9e-05;
  Matches  140;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy          5 CCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCACATCG 64
              ||  ||    | | |    |  |  | | || ||| || | |||   ||| | ||| |  |
Db        668 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 609

Qy         65 CCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGCGACC 124
              || ||  ||    || | || |||||  | ||  ||||||| || | |   || |  ||
Db        608 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 549

Qy        125 TCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCGGAC 184
              |  |  |  |||||  | || ||  |   |  | |  ||  |||||||  | || ||   |
Db        548 CCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCTCTCCCCCCCCCCCCCC 489

Qy        185 TCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCACCACC 244
              || || |||| || ||| | ||  |  || ||||  |||| ||||||   ||| || ||
Db        488 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCC 429

Qy        245 TC 246
              ||
Db        428 TC 427


RESULT 7
BI416575/c
LOCUS       BI416575                931 bp    mRNA    linear   EST 15-AUG-2001
DEFINITION  hasp001xg15f Heterobasidion annosum - Scots pine infection stage
            subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
            annosum cDNA clone hasp001xg15f, mRNA sequence.
ACCESSION   BI416575
VERSION     BI416575.1  GI:15187598
KEYWORDS    EST.
SOURCE      Pinus sylvestris/Heterobasidion annosum.
  ORGANISM  Pinus sylvestris/Heterobasidion annosum
            Eukaryota; mixed EST libraries.
REFERENCE   1  (bases 1 to 931)
  AUTHORS   Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
  TITLE     Expressed sequence tags of randomly selected cDNA clones from the
            interaction of the root rot fungus (Heterobasidion annosum) with
            seedling roots of Scots pine (Pinus sylvestris)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Fred O. Asiegbu
            Dept. of Forest Mycology & Pathology
            Swedish University of Agriculture, Box 7026,S-750 07, Uppsala,
            Sweden
            Tel: +46 18 67 15 98
            Fax: +46 18 30 92 45
            Email: Fred.Asiegbu@mykopat.slu.se
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .931
                     /organism="Pinus sylvestris/Heterobasidion annosum"
                     /db_xref="taxon:169015"
                     /clone="hasp001xg15f"
                     /clone_lib="Heterobasidion annosum - Scots pine infection
                     stage subtraction cDNA library (hasp)"
                     /dev_stage="Seedling roots of scots pine were infected for
                     6 days with H. annosum"
                     /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive
                     hybridization cDNA library was constructed from  scots
                     pine roots infected for 6-days with mycelia of
                     Heterobasidon  annosum (FP5)."
BASE COUNT        9 a      7 c    884 g     15 t     16 others
ORIGIN

  Query Match             32.3%;  Score 79.4;  DB 13;  Length 931;
  Best Local Similarity   57.9%;  Pred. No. 4.8e-05;
  Matches  140;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;
```

```
  Query Match           32.8%;  Score 80.6;  DB 12;  Length 688;
  Best Local Similarity  57.9%;  Pred. No. 3.1e-05;
  Matches  140;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy          5 CCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCACATCG 64
              ||  ||   | ||    ||||  | | || ||| || | |||    || | ||| |  |
Db        656 CCCCCCCNCCCCANCCCAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 597

Qy         65 CCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGCGACC 124
              |||||  ||    || | |||||||  | ||  |||||||  || | |   || |  ||
Db        596 CCACCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCC 537

Qy        125 TCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCGGAC 184
               |  |  | || ||| | || ||  |       | |  ||  | | ||| | || ||   |
Db        536 CCCCCCCCCACCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCC 477

Qy        185 TCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCACCACC 244
               || || |||| ||||||   ||  |  || ||||  |||| || ||| | |||||||
Db        476 CCCCCCCCCCCCCCACCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCACA 417

Qy        245 TC 246
               |
Db        416 NC 415
```

```
RESULT 4
BG067997
LOCUS       BG067997               1228 bp    mRNA    linear   EST 26-JAN-2001
DEFINITION  H3060F02-3 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
            H3060F02 3', mRNA sequence.
ACCESSION   BG067997
VERSION     BG067997.1  GI:12550566
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1228)
  AUTHORS   Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka
            ,T.S., Carter,M.G. and Ko,M.S.H.
  TITLE     Verification and initial annotation of NIA mouse 15K cDNA clone set
  JOURNAL   Unpublished (2001)
COMMENT     Other_ESTs: H3060F02-5
            Contact: George J. Kargul
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            This clone set has been freely distributed to the community. Please
            visit http://lgsun.grc.nia.nih.gov/cDNA/15k.html for details.
            Plate: H3060  row: F  column: 02
            Seq primer: -21M13 Forward
            High quality sequence stop: 1228
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .1228
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:H3060F02-3"
                     /db_xref="taxon:10090"
                     /clone="H3060F02"
                     /clone_lib="NIA Mouse 15K cDNA Clone Set"
                     /sex="Clones arrayed from a variety of cDNA libraries"
                     /dev_stage="Clones arrayed from a variety of cDNA
                     libraries"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
                     clone is among a rearrayed set of 15,247 clones from 11
                     embryo  cDNA libraries (including preimplantation stage
                     embryos from  unfertilized egg to blastocyst, embryonic
                     part of E7.5 embryos,  extraembryonic part of E7.5 embryos
                     , and E12.5 female mesonephros/gonad)  and one newborn
                     ovary cDNA library. Average insert size 1.5 kb. All
                     source libraries are cloned unidirectionally with Oligo(dT
                     )-Not primers.  References include: (1) Genome-wide
                     expression profiling of mid-gestation  placenta and embryo
                     using a 15,000 mouse developmental cDNA microarray,  2000,
                     Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)
                     Large-scale cDNA  analysis reveals phased gene expression
                     patterns during preimplantation  mouse develolpment, 2000,
                     Development, 127: 1737-1749; (3) Genome-wide  mapping of
                     unselected transcripts from extraembryonic tissue of
                     7.5-day  mouse embryos reveals enrichment in the t-complex
                     and under-representation  on the X chromosome, 1998, Hum
                     Mol Genet 7: 1967-1978."
BASE COUNT      157 a    858 c    103 g    108 t      2 others
ORIGIN
```

```
  Query Match           32.4%;  Score 79.8;  DB 12;  Length 1228;
  Best Local Similarity  58.0%;  Pred. No. 4e-05;
  Matches  141;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy          4 GCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCACATC 63
              |||  ||   | | |  ||  |  ||||||  || || | |||     |  | ||| |
Db        862 GCCCCCCCCCCCCCCCCACCCCCCCACACCCACCCCCCCCCCCCCCCCACCCCCCCCCCA 921

Qy         64 GCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGCGAC 123
               || ||  ||    |||| || |||||  | ||| |||||||| ||| | |    | | ||
Db        922 CCCCCCCCCCGCCCCCACCCCCCCCCCCCCCCCCACCCCCCCCCACCCCCCCGCACACCAC 981

Qy        124 CTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCGGA 183
               | |  |  | | ||| | || ||  |   |  | |  ||  | | ||  | |||||
Db        982 CCCCCCCACACCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCACCCCCACCCCC 1041

Qy        184 CTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCACCAC 243
               | || || |||| |||||| ||||  |  || ||||  |||| || ||| | ||| || |
Db       1042 CCCCCCCCCCCCCCCACCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 1101

Qy        244 CTC 246
               | |
Db       1102 CCC 1104
```

```
RESULT 5
CNS01MZJ
LOCUS       CNS01MZJ               954 bp    DNA     linear   GSS 14-JUN-2001
DEFINITION  Anopheles gambiae GSS T7 end of clone 22N11 of NotreDame1 library
            from strain PEST of Anopheles gambiae (African malaria mosquito),
            genomic survey sequence.
ACCESSION   AL151584
VERSION     AL151584.1  GI:7012063
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 954)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
REFERENCE   2  (bases 1 to 954)
  AUTHORS   Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
            Roux, Paris 75015, France
COMMENT     This clone is from an A. gambiae BAC library provided by F.H.
            Collins and sequenced by Genoscope in collaboration with the
            Laboratory of Biochem. and Biol. Molec. of Insects, Institut
            Pasteur.
FEATURES             Location/Qualifiers
     source          1. .954
                     /organism="Anopheles gambiae"
```

```
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="PC_0028_A2_A08_MR"
                     /clone_lib="Sea urchin primary mesenchyme cell cDNA
                     library"
                     /tissue_type="embryo"
                     /cell_type="primary mesenchyme cells"
                     /lab_host="E.coli"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; oligo
                     dt priming from poly A+ RNA, directionally cloned"
BASE COUNT        7 a      0 c    667 g    117 t     34 others
ORIGIN

  Query Match             34.6%;  Score 85.2;  DB 12;  Length 825;
  Best Local Similarity   58.3%;  Pred. No. 4.7e-06;
  Matches  141;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy          5 CCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCACATCG 64
              ||  ||  || | |   |  |  | |||| ||| || | |||    || | |||||  |
Db        411 CCCCCCCCACCCCCCNNCNCCCCCCCCACCCCCCCCCCCCCCCCCACCCCCCCCCCCACCCCC 352

Qy         65 CCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGCGACC 124
              || ||  ||     | | || |||||  | ||  || ||||  ||    |   || | |||
Db        351 CCCCCCCCCACCCCNCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCNCCCCCCCCCCACC 292

Qy        125 TCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCGGAC 184
               |  |  | || ||  |  | ||  | | | || | |||  | | ||| | || ||    |
Db        291 CCCCCCCCCACCCCNACCACCCCCCCCACCCACCCCACCCCCACCCCCACCCCCCCCCCC 232

Qy        185 TCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCACCACC 244
               || || ||||||| ||| | ||  |   | |||| |||||||| | | |||||||| ||
Db        231 CCCCCCCCCCCACCNCCCCCCCCCCCCCCCACCCCCCCACCCCACCCCACCCACCCACCCC 172

Qy        245 TC 246
               |
Db        171 CC 170


RESULT 2
AG047947/c
LOCUS       AG047947                 855 bp    DNA     linear   GSS 02-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-027J24.F, genomic survey sequence.
ACCESSION   AG047947
VERSION     AG047947.1  GI:16584839
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-027J24.F.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 855)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .855
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-027J24.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT       75 a     11 c    533 g    175 t     61 others
ORIGIN

  Query Match             33.2%;  Score 81.6;  DB 17;  Length 855;
  Best Local Similarity   57.5%;  Pred. No. 2e-05;
  Matches  138;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy          5 CCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCACATCG 64
              ||  |||||||||  ||| ||| | | || ||| || | ||      |   ||||||
Db        493 CCCACCAGACACAAAAACCACACCCCCCCCCCCCCCCCCCANCCCCACNTCCCACCCAC 434

Qy         65 CCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGCGACC 124
              ||||| | |    |  |||| |||||  |||| || |||||  || | |    ||    ||
Db        433 CCACCATACCCCACACCACCCCCCCCCCCCACCTTCTCCCCCCCCCCCCCCCCCCANCCCC 374

Qy        125 TCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCGGAC 184
               |  |  |  | ||| |||| ||  |   | || |  ||  | | ||| | || ||    |
Db        373 CCCCCCCCCCCCCCCCCACCCCCCACCCCCCACCCCCCCACCCCCCCCCCCCCCCCCCC 314

Qy        185 TCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCACCACC 244
               || || ||||    ||| |  |    ||| |||| |||| |||||| |   | || |
Db        313 CCCCCCCCCCCCNACCCCCCCCACCCACACCCCCCCCNCCCCCCCACCCCCCNNCCCCCNC 254


RESULT 3
BG786340/c
LOCUS       BG786340                 688 bp    mRNA    linear   EST 20-MAY-2001
DEFINITION  SEAUMC006297 Sea urchin primary mesenchyme cell cDNA library
            Strongylocentrotus purpuratus cDNA clone PC_0028_A2_H09_MR 5', mRNA
            sequence.
ACCESSION   BG786340
VERSION     BG786340.1  GI:14157353
KEYWORDS    EST.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 688)
  AUTHORS   Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and
            Ettensohn,C.A.
  TITLE     A large scale analysis of mRNAs expressed by primary mesenchyme
            cells of the sea urchin embryo
  JOURNAL   Development 128 (13), 2615-2627 (2001)
  MEDLINE   21384984
COMMENT     Contact: Ettensohn CA
            Dept. Biol. Sci.
            Carnegie Mellon University
            4400 Fifth Avenue, Pittsburgh, PA 15213, USA
            Tel: +1 412 268 5849
            Email: ettensohn@andrew.cmu.edu.
FEATURES             Location/Qualifiers
     source          1. .688
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="PC_0028_A2_H09_MR"
                     /clone_lib="Sea urchin primary mesenchyme cell cDNA
                     library"
                     /tissue_type="embryo"
                     /cell_type="primary mesenchyme cells"
                     /lab_host="E.coli"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; oligo
                     dt priming from poly A+ RNA, directionally cloned"
BASE COUNT        9 a      0 c    596 g     53 t     30 others
ORIGIN
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 315.347 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_11413_11658
Perfect score: 246
Sequence: 1 ctggccggccagacacacga.....ccaccctcacccaccacctc 246

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	85.2	34.6	825	12	BG786255	BG786255 SEAUMC006
c	2	81.6	33.2	855	17	AG047947	AG047947 Pan trogl
c	3	80.6	32.8	688	12	BG786340	BG786340 SEAUMC006
	4	79.8	32.4	1228	12	BG067997	BG067997 H3060F02-
	5	79.6	32.4	954	17	CNS01MZJ	AL151584 Anopheles
c	6	79.4	32.3	803	13	BI416474	BI416474 hasp001xa
c	7	79.4	32.3	931	13	BI416575	BI416575 hasp001xg
	8	79.2	32.2	389	17	CNS04FIQ	AL288395 Tetraodon
c	9	79	32.1	646	12	BG786282	BG786282 SEAUMC006
c	10	79	32.1	744	17	AG153752	AG153752 Pan trogl
c	11	79	32.1	912	17	AG136783	AG136783 Pan trogl
	12	78.8	32.0	375	12	BG519974	BG519974 ps07c09.y
	13	78.8	32.0	585	10	AW335294	AW335294 S45E8 AGS
	14	78.8	32.0	848	17	CNS0118A	AL099940 Drosophil
c	15	78.8	32.0	913	17	AZ687957	AZ687957 ENTIJ61TR
c	16	78.6	32.0	806	13	BI416659	BI416659 hasp001xm
c	17	78.6	32.0	1087	17	AG043580	AG043580 Pan trogl
	18	78.4	31.9	1101	17	CNS0181N	AL108773 Drosophil
	19	78.2	31.8	1303	14	BM908878	BM908878 AGENCOURT
c	20	78	31.7	715	9	AL570062	AL570062 AL570062
c	21	77.8	31.6	738	17	AZ199465	AZ199465 SP_1039_B
c	22	77.8	31.6	818	17	AG060534	AG060534 Pan trogl
c	23	77.8	31.6	960	17	CNS0048W	AL051640 Drosophil
	24	77.6	31.5	586	17	CNS01UD1	AL167662 Tetraodon
c	25	77.6	31.5	669	17	CNS032YM	AL225463 Tetraodon
c	26	77.6	31.5	736	17	AZ197707	AZ197707 SP_1035_B
c	27	77.6	31.5	929	17	AZ687353	AZ687353 ENTJT25TF
c	28	77.4	31.5	261	17	CNS04KHD	AL294826 Tetraodon
c	29	77.4	31.5	637	17	AG077880	AG077880 Pan trogl
c	30	77.4	31.5	823	17	AZ187132	AZ187132 SP_1008_B
	31	77.4	31.5	1088	13	BM547778	BM547778 AGENCOURT
c	32	77.4	31.5	1166	12	BF256751	BF256751 HVSMEf001
	33	77.2	31.4	529	10	AW050654	AW050654 wz13g11.x
c	34	77.2	31.4	559	17	CNS04CPE	AL284747 Tetraodon
	35	77.2	31.4	593	10	AW320968	AW320968 un17h12.x
	36	77.2	31.4	638	17	CNS046XY	AL277279 Tetraodon
c	37	77.2	31.4	740	17	AZ195718	AZ195718 SP_1030_B
c	38	77.2	31.4	812	9	AL537183	AL537183 AL537183
c	39	77.2	31.4	843	17	AZ195427	AZ195427 SP_1030_A
c	40	77.2	31.4	903	17	BH136778	BH136778 ENTPA49TF
	41	77.2	31.4	942	9	AL536132	AL536132 AL536132
c	42	77.2	31.4	945	17	AG117065	AG117065 Pan trogl
	43	77	31.3	424	10	AW778904	AW778904 ho16f12.x
	44	77	31.3	459	9	AI285778	AI285778 qu56b02.x
	45	77	31.3	928	9	AL561014	AL561014 AL561014

ALIGNMENTS

RESULT 1
BG786255/c
LOCUS BG786255 825 bp mRNA linear EST 20-MAY-2001
DEFINITION SEAUMC006212 Sea urchin primary mesenchyme cell cDNA library Strongylocentrotus purpuratus cDNA clone PC_0028_A2_A08_MR 5', mRNA sequence.
ACCESSION BG786255
VERSION BG786255.1 GI:14157268
KEYWORDS EST.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.
REFERENCE 1 (bases 1 to 825)
AUTHORS Zhu,X., Mahairas,G., Illies,M.R., Cameron,R.A., Davidson,E.H. and Ettensohn,C.A.
TITLE A large scale analysis of mRNAs expressed by primary mesenchyme cells of the sea urchin embryo
JOURNAL Development 128 (13), 2615-2627 (2001)
MEDLINE 21384984
COMMENT Contact: Ettensohn CA
Dept. Biol. Sci.
Carnegie Mellon University
4400 Fifth Avenue, Pittsburgh, PA 15213, USA
Tel: +1 412 268 5849
Email: ettensohn@andrew.cmu.edu.
FEATURES Location/Qualifiers
source 1. .825

```
RESULT 14
US-09-861-289-34
; Sequence 34, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-34

  Query Match             45.2%;  Score 111.2;  DB 10;  Length 4689;
  Best Local Similarity   70.3%;  Pred. No. 1.1e-18;
  Matches  149;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy         35 CCCTCCTCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACACCACCCCCGACA 94
              | |||||| |||| |||||  |||||   ||  |||| || |||||  || ||||||||
Db       4211 CGCTCCTCACCCTCGTCCGTACCCACGCGGCGGCCGTACTCGGCCATTCCTCCCCCGACC 4270

Qy         95 CCATCCCCCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCACCGCCGTCGAAC 154
              |  |||||| ||| || |||  ||| ||||||||||||||| || || |||||  | |
Db       4271 GGGTGGCCCCCGGCCGTGCCTTCACCGAGCTCGGCTTCGACTCGCTGACGGCCGTGCAGC 4330

Qy        155 TACGCAACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCTCGCCTTCGACC 214
              | ||||||| |||||||         |||   | | |||||| |||||   | ||||||||
Db       4331 TCCGCAACCAGCTCTCCACGGTGGTCGGCAACAGGCTCCCCGCCACCACGGTCTTCGACC 4390

Qy        215 ACCCCAACCCCACCACCCTCACCCACCACCTC 246
              |||| |  ||| || | ||| ||   ||||||
Db       4391 ACCCGACGCCCGCCGCACTCGCCGCGCACCTC 4422


RESULT 15
US-09-860-846-30
; Sequence 30, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-30

  Query Match             44.7%;  Score 110;  DB 9;  Length 13842;
  Best Local Similarity   68.5%;  Pred. No. 1.9e-18;
  Matches  152;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy         24 ACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACAC 83
              || |||   |  || ||||  | |||  |  | |  |||||  || || |||||| |
Db      13212 ACGGCAGCGCCTGCTGCTCGAGTTCGTCGTCGGCGAGGTCGCCGAAGTACTCGGCCACGC 13271

Qy         84 CACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCAC 143
              |  |  ||  ||| | |||| || |  |||| ||||||||||||||||||||||| ||
Db      13272 CCGCGGTCACCGGATCGACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGAC 13331

Qy        144 CGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCT 203
              |||||||||||| |||||||||||  |  | ||   || |||  ||||||  | |||||
Db      13332 CGCCGTCGAACTCCGCAACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGCGACCCT 13391

Qy        204 CGCCTTCGACCACCCCAACCCCACCACCCTCACCCACCACCT 245
              | ||||||||||||  | |||  |  | |||  ||  ||||||
Db      13392 GGTCTTCGACCACCCAAGCCCGGCGGCACTCGCCTCCCACCT 13433


Search completed: June 18, 2003, 05:08:47
Job time : 30.1628 secs
```

```
Db      29173 CCCTGCGCGACCTCGTCCGCGGCCACGTCGCCCTGGTGCTCGGACACAGTGGCCCCGAGG 29232

Qy         95 CCATCCCCCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCACCGCCGTCGAAC 154
              |||||  | |||| |  || ||||| ||| |||| |||||||||||| ||||| |||||||
Db      29233 CCATCGACGCCGAACAGGCCTTCCGGGACATCGGTTTCGACTCCCTGACCGCAGTCGAAC 29292

Qy        155 TACGCAACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCTCGCCTTCGACC 214
              |  | |||||||||    |  |   ||||| |||||| | |||  ||| ||||  ||||||
Db      29293 TCAGAAACCGGCTGAACGCCGAGACCGGCCTCCGCTTGCCCGGCACGCTCGTGTTCGACT 29352

Qy        215 ACCCCAACCCCACCACCCTCACCCACCACCT 245
              |||||||||| | | | ||| || | |||||
Db      29353 ACCCCAACCCGAGCGCGCTCGCCGATCACCT 29383


RESULT 11
US-09-860-846-34
; Sequence 34, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-34

  Query Match             45.2%;  Score 111.2;  DB 9;  Length 4689;
  Best Local Similarity   70.3%;  Pred. No. 1.1e-18;
  Matches  149;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy         35 CCCTCCTCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACACCACCCCCGACA 94
              | |||||| |||| |||||  |||||   ||  |||| || |||||  || |||||||||
Db       4211 CGCTCCTCACCCTCGTCCGTACCCACGCGGCGGCCGTACTCGGCCATTCCTCCCCCGACC 4270

Qy         95 CCATCCCCCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCACCGCCGTCGAAC 154
                 |  |||||| ||| || |||  ||| ||||||||||||||| || || |||||  | |
Db       4271 GGGTGGCCCCCGGCCGTGCCTTCACCGAGCTCGGCTTCGACTCGCTGACGGCCGTGCAGC 4330

Qy        155 TACGCAACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCTCGCCTTCGACC 214
              | |||||||  |||||||         |||   | | ||||||  |||||   | ||||||||
Db       4331 TCCGCAACCAGCTCTCCACGGTGGTCGGCAACAGGCTCCCCGCCACCACGGTCTTCGACC 4390

Qy        215 ACCCCAACCCCACCACCCTCACCCACCACCTC 246
              ||||  |  |||  || |  ||| ||   ||||||
Db       4391 ACCCGACGCCCGCCGCACTCGCCGCGCACCTC 4422


RESULT 12
US-09-988-384B-34
; Sequence 34, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-34

  Query Match             45.2%;  Score 111.2;  DB 9;  Length 4689;
  Best Local Similarity   70.3%;  Pred. No. 1.1e-18;
  Matches  149;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy         35 CCCTCCTCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACACCACCCCCGACA 94
              | |||||| |||| |||||  |||||   ||  |||| || |||||  || |||||||||
Db       4211 CGCTCCTCACCCTCGTCCGTACCCACGCGGCGGCCGTACTCGGCCATTCCTCCCCCGACC 4270

Qy         95 CCATCCCCCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCACCGCCGTCGAAC 154
                 |  |||||| ||| || |||  ||| ||||||||||||||| || || |||||  | |
Db       4271 GGGTGGCCCCCGGCCGTGCCTTCACCGAGCTCGGCTTCGACTCGCTGACGGCCGTGCAGC 4330

Qy        155 TACGCAACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCTCGCCTTCGACC 214
              | |||||||  |||||||         |||   | | ||||||  |||||   | ||||||||
Db       4331 TCCGCAACCAGCTCTCCACGGTGGTCGGCAACAGGCTCCCCGCCACCACGGTCTTCGACC 4390

Qy        215 ACCCCAACCCCACCACCCTCACCCACCACCTC 246
              ||||  |  |||  || |  ||| ||   ||||||
Db       4391 ACCCGACGCCCGCCGCACTCGCCGCGCACCTC 4422


RESULT 13
US-09-836-821-34
; Sequence 34, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-34

  Query Match             45.2%;  Score 111.2;  DB 9;  Length 4689;
  Best Local Similarity   70.3%;  Pred. No. 1.1e-18;
  Matches  149;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy         35 CCCTCCTCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACACCACCCCCGACA 94
              | |||||| |||| |||||  |||||   ||  |||| || |||||  || |||||||||
Db       4211 CGCTCCTCACCCTCGTCCGTACCCACGCGGCGGCCGTACTCGGCCATTCCTCCCCCGACC 4270

Qy         95 CCATCCCCCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCACCGCCGTCGAAC 154
                 |  |||||| ||| || |||  ||| ||||||||||||||| || || |||||  | |
Db       4271 GGGTGGCCCCCGGCCGTGCCTTCACCGAGCTCGGCTTCGACTCGCTGACGGCCGTGCAGC 4330

Qy        155 TACGCAACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCTCGCCTTCGACC 214
              | |||||||  |||||||         |||   | | ||||||  |||||   | ||||||||
Db       4331 TCCGCAACCAGCTCTCCACGGTGGTCGGCAACAGGCTCCCCGCCACCACGGTCTTCGACC 4390

Qy        215 ACCCCAACCCCACCACCCTCACCCACCACCTC 246
              ||||  |  |||  || |  ||| ||   ||||||
Db       4391 ACCCGACGCCCGCCGCACTCGCCGCGCACCTC 4422
```

```
Db      26605 CACCT 26609


RESULT 8
US-09-988-384B-5
; Sequence 5, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 5
;   LENGTH: 37948
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-5

  Query Match             48.2%;  Score 118.6;  DB 9;  Length 37948;
  Best Local Similarity   67.8%;  Pred. No. 1.4e-20;
  Matches  166;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||||    |||||  ||  | | |  | |   ||| |  | ||| |||||  | |||
Db      27535 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 27594

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 120
                | || |||||||||| |||  ||||||   |   |   ||  || || || ||||||
Db      27595 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 27654

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||  ||| ||||||||| ||||||||||||||||||| ||||||| |||  |  | | |||
Db      27655 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 27714

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || || || || ||  |||| || | ||||||||||||||| |||        ||| ||
Db      27715 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 27774

Qy        241 CACCT 245
              |||||
Db      27775 CACCT 27779


RESULT 9
US-09-793-708-19
; Sequence 19, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-19

  Query Match             48.2%;  Score 118.6;  DB 9;  Length 38506;
  Best Local Similarity   67.8%;  Pred. No. 1.4e-20;
  Matches  166;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||||    |||||  ||  | | |  | |   ||| |  | ||| |||||  | |||
Db      24507 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 24566

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 120
                | || |||||||||| |||  ||||||   |   |   ||  || || || ||||||
Db      24567 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 24626

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||  ||| ||||||||| ||||||||||||||||||| ||||||| |||  |  | | |||
Db      24627 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 24686

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || || || || ||  |||| || | ||||||||||||||| |||        ||| ||
Db      24687 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 24746

Qy        241 CACCT 245
              |||||
Db      24747 CACCT 24751


RESULT 10
US-09-808-880-1
; Sequence 1, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT:  Shah, Sanjay Krishnakant
;  APPLICANT:  McDaniel, Robert
;  APPLICANT:  Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/808,880
;  CURRENT FILING DATE:  2001-03-14
;  PRIOR APPLICATION NUMBER: US/09/428,517
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/120,254
;  PRIOR FILING DATE: 1999-02-16
;  PRIOR APPLICATION NUMBER: 60/106,100
;  PRIOR FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 50937
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artifical Sequence: Recombinant DNA
US-09-808-880-1

  Query Match             48.0%;  Score 118.2;  DB 9;  Length 50937;
  Best Local Similarity   72.5%;  Pred. No. 1.7e-20;
  Matches  153;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy         35 CCCTCCTCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACACCACCCCCGACA 94
              |||| | || ||| ||||||  |||| |||||   || || || ||||    ||||||
```

```
Qy        241 CACCT 245
              |||||
Db      10918 CACCT 10922


RESULT 5
US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-5

  Query Match             48.2%;  Score 118.6;  DB 9;  Length 36778;
  Best Local Similarity   67.8%;  Pred. No. 1.4e-20;
  Matches  166;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||||    |||||  ||  | | |  | |   ||| |  | ||| |||||  | |||
Db      26365 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 26424

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 120
               | || ||||||||||| ||| ||||||   |     |   ||  || || || ||||||
Db      26425 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 26484

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||  ||| |||||||| |||||||||||||||||| |||||||| |||  |  | | |||
Db      26485 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 26544

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || || || || ||  |||| || | |||||||||||||| |||       ||| ||
Db      26545 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 26604

Qy        241 CACCT 245
              |||||
Db      26605 CACCT 26609


RESULT 6
US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-5

  Query Match             48.2%;  Score 118.6;  DB 9;  Length 36778;
  Best Local Similarity   67.8%;  Pred. No. 1.4e-20;
  Matches  166;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||||    |||||  ||  | | |  | |   ||| |  | ||| |||||  | |||
Db      26365 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 26424

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 120
               | || ||||||||||| ||| ||||||   |     |   ||  || || || ||||||
Db      26425 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 26484

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||  ||| |||||||| |||||||||||||||||| |||||||| |||  |  | | |||
Db      26485 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 26544

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || || || || ||  |||| || | |||||||||||||| |||       ||| ||
Db      26545 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 26604

Qy        241 CACCT 245
              |||||
Db      26605 CACCT 26609


RESULT 7
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-5

  Query Match             48.2%;  Score 118.6;  DB 10;  Length 36778;
  Best Local Similarity   67.8%;  Pred. No. 1.4e-20;
  Matches  166;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||||    |||||  ||  | | |  | |   ||| |  | ||| |||||  | |||
Db      26365 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 26424

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 120
               | || ||||||||||| ||| ||||||   |     |   ||  || || || ||||||
Db      26425 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 26484

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||  ||| |||||||| |||||||||||||||||| |||||||| |||  |  | | |||
Db      26485 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 26544

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || || || || ||  |||| || | |||||||||||||| |||       ||| ||
Db      26545 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 26604

Qy        241 CACCT 245
              |||||
```

```
Qy        241 CACCT 245
              |||||
Db      10918 CACCT 10922


RESULT 2
US-09-988-384B-32
; Sequence 32, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-32

  Query Match             48.2%;  Score 118.6;  DB 9;  Length 11220;
  Best Local Similarity   67.8%;  Pred. No. 1.7e-20;
  Matches  166;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||||    |||||  ||  | | |  | |   ||| |  | ||| |||||  | |||
Db      10678 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 10737

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 120
               | || ||||||||||| |||  ||||||    |     |  ||  || || || ||||||
Db      10738 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 10797

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||  ||| |||||||| |||||||||||||||||| ||||||| |||  |  | | |||
Db      10798 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 10857

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              ||.|| || || ||  |||| || | |||||||||||||| |||        ||| ||
Db      10858 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 10917

Qy        241 CACCT 245
              |||||
Db      10918 CACCT 10922


RESULT 3
US-09-836-821-32
; Sequence 32, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-32

  Query Match             48.2%;  Score 118.6;  DB 9;  Length 11220;
  Best Local Similarity   67.8%;  Pred. No. 1.7e-20;
  Matches  166;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||||    |||||  ||  | | |  | |   ||| |  | ||| |||||  | |||
Db      10678 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 10737

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 120
               | || ||||||||||| |||  ||||||    |     |  ||  || || || ||||||
Db      10738 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 10797

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||  ||| |||||||| |||||||||||||||||| ||||||| |||  |  | | |||
Db      10798 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 10857

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || || || || ||  |||| || | |||||||||||||| |||        ||| ||
Db      10858 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 10917

Qy        241 CACCT 245
              |||||
Db      10918 CACCT 10922


RESULT 4
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-32

  Query Match             48.2%;. Score 118.6;  DB 10;  Length 11220;
  Best Local Similarity   67.8%;  Pred. No. 1.7e-20;
  Matches  166;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||||    |||||  ||  | | |  | |   ||| |  | ||| |||||  | |||
Db      10678 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 10737

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 120
               | || ||||||||||| |||  ||||||    |     |  ||  || || || ||||||
Db      10738 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 10797

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||  ||| |||||||| |||||||||||||||||| ||||||| |||  |  | | |||
Db      10798 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 10857

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || || || || ||  |||| || | |||||||||||||| |||        ||| ||
Db      10858 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 10917
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 17:56:34 ; Search time 29.0718 Seconds
(without alignments)
12253.192 Million cell updates/sec

Title: US-09-914-286-1_COPY_11413_11658
Perfect score: 246
Sequence: 1 ctggccggccagacacacga.....ccaccctcacccaccacctc 246

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	118.6	48.2	11220	9	US-09-860-846-32	Sequence 32, Appl
	2	118.6	48.2	11220	9	US-09-988-384B-32	Sequence 32, Appl
	3	118.6	48.2	11220	9	US-09-836-821-32	Sequence 32, Appl
	4	118.6	48.2	11220	10	US-09-861-289-32	Sequence 32, Appl
	5	118.6	48.2	36778	9	US-09-860-846-5	Sequence 5, Appli
	6	118.6	48.2	36778	9	US-09-836-821-5	Sequence 5, Appli
	7	118.6	48.2	36778	10	US-09-861-289-5	Sequence 5, Appli
	8	118.6	48.2	37948	9	US-09-988-384B-5	Sequence 5, Appli
	9	118.6	48.2	38506	9	US-09-793-708-19	Sequence 19, Appl
	10	118.2	48.0	50937	9	US-09-808-880-1	Sequence 1, Appli
	11	111.2	45.2	4689	9	US-09-860-846-34	Sequence 34, Appl
	12	111.2	45.2	4689	9	US-09-988-384B-34	Sequence 34, Appl
	13	111.2	45.2	4689	9	US-09-836-821-34	Sequence 34, Appl
	14	111.2	45.2	4689	10	US-09-861-289-34	Sequence 34, Appl
	15	110	44.7	13842	9	US-09-860-846-30	Sequence 30, Appl
	16	110	44.7	13842	9	US-09-988-384B-30	Sequence 30, Appl
	17	110	44.7	13842	9	US-09-836-821-30	Sequence 30, Appl
	18	110	44.7	13842	10	US-09-861-289-30	Sequence 30, Appl
	19	103.6	42.1	15872	9	US-09-860-846-1	Sequence 1, Appli
	20	103.6	42.1	15872	9	US-09-988-384B-1	Sequence 1, Appli
	21	103.6	42.1	15872	9	US-09-836-821-1	Sequence 1, Appli
	22	103.6	42.1	15872	10	US-09-861-289-1	Sequence 1, Appli
	23	95.6	38.9	4041	9	US-09-860-846-36	Sequence 36, Appl
	24	95.6	38.9	4041	9	US-09-988-384B-36	Sequence 36, Appl
	25	95.6	38.9	4041	9	US-09-836-821-36	Sequence 36, Appl
	26	95.6	38.9	4041	10	US-09-861-289-36	Sequence 36, Appl
	27	86	35.0	4851	9	US-09-712-363-116	Sequence 116, App
	28	75.4	30.7	4209	9	US-09-712-363-20	Sequence 20, Appl
c	29	73.2	29.8	520	9	US-10-184-644-332	Sequence 332, App
c	30	73.2	29.8	520	9	US-10-184-634-332	Sequence 332, App
c	31	73	29.7	12733	9	US-10-032-393-47	Sequence 47, Appl
c	32	73	29.7	12739	9	US-10-032-393-8	Sequence 8, Appli
	33	69.4	28.2	68750	9	US-10-014-717-1	Sequence 1, Appli
	34	67	27.2	925	9	US-09-735-056-1	Sequence 1, Appli
	35	67	27.2	152331	9	US-10-095-407-16	Sequence 16, Appl
	36	61.8	25.1	1064	10	US-09-804-682-29	Sequence 29, Appl
	37	60.8	24.7	504	10	US-09-878-574-4330	Sequence 4330, Ap
	38	59.4	24.1	1065	10	US-09-804-682-33	Sequence 33, Appl
	39	58.8	23.9	266	10	US-09-960-352-3882	Sequence 3882, Ap
c	40	57.6	23.4	2163	9	US-10-198-846-12766	Sequence 12766, A
	41	57	23.2	437	10	US-09-960-352-12835	Sequence 12835, A
c	42	56.6	23.0	931	9	US-10-198-846-8585	Sequence 8585, Ap
	43	56.2	22.8	987	10	US-09-804-682-20	Sequence 20, Appl
	44	55.6	22.6	1194	10	US-09-804-682-7	Sequence 7, Appli
	45	55.6	22.6	3453	9	US-10-101-464A-861	Sequence 861, App

ALIGNMENTS

RESULT 1
US-09-860-846-32
; Sequence 32, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-32

Query Match 48.2%; Score 118.6; DB 9; Length 11220;
Best Local Similarity 67.8%; Pred. No. 1.7e-20;
Matches 166; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

```
Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||||    ||||| ||  | | |  | |   ||| |  | ||| |||||  | |||
Db      10678 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 10737

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
               | || |||||||||| ||| ||||||   |     |   ||  || || || ||||||
Db      10738 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 10797

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||  ||| |||||||| |||||||||||||||||| |||||||| |||  |  | | |||
Db      10798 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 10857

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || || || || ||  |||| || | ||||||||||||||| |||       ||| ||
Db      10858 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 10917
```

THIS PAGE BLANK (USPTO)

```
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  80161 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-036-987A-1

  Query Match             43.7%;  Score 107.4;  DB 3;  Length 80161;
  Best Local Similarity   64.9%;  Pred. No. 4.3e-13;
  Matches  159;  Conservative    0;  Mismatches   86;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || || |  | | |  | |  ||||  | |  |   ||  | |  |||||  || ||||
Db      39790 CTCGCGGAGCTGCCGGAAGCCCAACGACGCCACGAACTGGTGGATCTGGTGTGCGCCCAG 39849

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
               | || |||||  || || |||  ||  | |||      |||  |||||  || |||||||||
Db      39850 GTGGCAACCGTGCTCGGGCACGGCAGTCGCGAGGAAGTCCAGCCCGAGCGGGCGTTCCGC 39909

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |  |||||  |||||||||||||  || || || || |||||  ||  |  ||  | |||||
Db      39910 GCGCTCGGGTTCGACTCCCTCATGGCGGTGGATCTGCGCAATCGTTTGACCACCGCCACC 39969

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              ||  | |||||  || |||||   || |||||||  |||| || ||  || || |  ||
Db      39970 GGGTTGCGCCTGCCGACCACAACCGTCTTCGACTACCCGAATCCGGCCGCCTTGGCCGCT 40029

Qy        241 CACCT 245
              |||||
Db      40030 CACCT 40034


Search completed: June 18, 2003, 01:08:30
Job time : 26.6565 secs
```

```
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  97..1482
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacyl ACP synthase of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1693..2670
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3406..3921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4171..4428
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 3"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4471..10722
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  4471..5847
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynhase domain of module"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6054..7026
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  7165..9216
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  dehydratase and enoylreductase domains m"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  9433..9984
;       OTHER INFORMATION:  /function= "approximate span
;       OTHER INFORMATION:  beta-ketoreductase of module 4"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10225..10483
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 4"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  10723..20235
;       OTHER INFORMATION:  /codon_start= 10723
;       OTHER INFORMATION:  /function= "gene =eryA"
;       OTHER INFORMATION:  /product= "orf3 encoding modules 5 & 6
;       OTHER INFORMATION:  6-deoxyerythronolide B formatio"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10723..15165
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10831..12174
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  12379..13350
;       OTHER INFORMATION:  /function= "approximatr span of
;       OTHER INFORMATION:  acyltransferase domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14062..14610
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14857..15114
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15166..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15172..16569
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  16768..17721
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  18379..18921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19149..19398
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19492..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  thioesterase domain of module 6"
US-08-439-009A-3

  Query Match             44.2%;  Score 108.8;  DB 3;  Length 20235;
  Best Local Similarity   70.2%;  Pred. No. 2.3e-13;
  Matches  146;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy         36 CCTCCTCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACACCACCCCCGACAC 95
              ||||  |  |||| ||||| || |  ||||| | || || ||||||  |  |  ||   |
Db      14892 CCTCGCCCACCTGATCCGCGCCGAGGTCGCCGCGGTGCTCGGCCACGGCGACGACGCGGC 14951

Qy         96 CATCCCCCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCACCGCCGTCGAACT 155
              |||  || |||||||| ||||||||||||||||||||||||| | ||||||||||| ||
Db      14952 GATCGACCGCGACCGCGCCTTCCGCGACCTCGGCTTCGACTCCATGACCGCCGTCGACCT 15011

Qy        156 ACGCAACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCTCGCCTTCGACCA 215
              || |||||||| ||      |||||  | ||     || | ||  | | |||||||||
Db      15012 GCGGAACCGGCTCGCCGCGGTGACCGGGGTGCGGGAAGCCGCGACGGTGGTCTTCGACCA 15071

Qy        216 CCCCAACCCCACCACCCTCACCCACCAC 243
              ||| | |  ||||   ||| || |||||
Db      15072 CCCGACCATCACCCGGCTCGCCGACCAC 15099


RESULT 15
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
```

```
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10225..10483
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 4"
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  10723..20235
;       OTHER INFORMATION:  /codon_start= 10723
;       OTHER INFORMATION:  /function= "gene =eryA"
;       OTHER INFORMATION:  /product= "orf3 encoding modules 5 & 6
;       OTHER INFORMATION:  6-deoxyerythronolide B formatio"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10723..15165
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10831..12174
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  12379..13350
;       OTHER INFORMATION:  /function= "approximatr span of
;       OTHER INFORMATION:  acyltransferase domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14062..14610
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  14857..15114
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 5"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15166..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  15172..16569
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACPsynthase domain of modul"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  16768..17721
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  18379..18921
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19149..19398
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 6"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  19492..20235
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  thioesterase domain of module 6"
US-07-642-734C-3

  Query Match           44.2%;  Score 108.8;  DB 1;  Length 20235;
  Best Local Similarity  70.2%;  Pred. No. 2.3e-13;
  Matches  146;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy         36 CCTCCTCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACACCACCCCCGACAC 95
              ||||  |  |||| ||||| || |  ||||| | || || ||||||  |  |  ||   |
Db      14892 CCTCGCCCACCTGATCCGCGCCGAGGTCGCCGCGGTGCTCGGCCACGGCGACGACGCGGC 14951

Qy         96 CATCCCCCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCACCGCCGTCGAACT 155
              |||  || |||||||||| ||||||||||||||||||||||||| | ||||||||||| ||
Db      14952 GATCGACCGCGACCGCGCCTTCCGCGACCTCGGCTTCGACTCCATGACCGCCGTCGACCT 15011

Qy        156 ACGCAACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCTCGCCTTCGACCA 215
              || |||||||||| ||      |||||  | ||     || | ||  | | |||||||||
Db      15012 GCGGAACCGGCTCGCCGCGGTGACCGGGGTGCGGGAAGCCGCGACGGTGGTCTTCGACCA 15071

Qy        216 CCCCAACCCCACCACCCTCACCCACCAC 243
              ||| | |  ||||   ||| || |||||
Db      15072 CCCGACCATCACCCGGCTCGCCGACCAC 15099


RESULT 14
US-08-439-009A-3
; Sequence 3, Application US/08439009A
; Patent No. 6004787
;  GENERAL INFORMATION:
;    APPLICANT:  Donadio, S
;    APPLICANT:  Katz, L
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Method of Directing Biosynthesis of
;    TITLE OF INVENTION:  Specific Polyketides
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steven F. Weinstock
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,009A
;      FILING DATE:  11-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Casuto, Dianne
;      REGISTRATION NUMBER:  40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847-938-3137
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20235 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 238
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  19..10722
;      OTHER INFORMATION:  /codon_start= 19
;      OTHER INFORMATION:  /function= "gene eryA"
;      OTHER INFORMATION:  /product= "eryA ORF2 encoding modules 3 & 4 for
;      OTHER INFORMATION:  6-deoxyerythronolide B"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19..4470
```

```
RESULT 12
US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-30

  Query Match             44.7%;  Score 110;  DB 4;  Length 13842;
  Best Local Similarity   68.5%;  Pred. No. 1.3e-13;
  Matches  152;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy         24 ACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACAC 83
              || |||    |   || ||||   | |||  |  | |  |||||   || || |||||| |
Db      13212 ACGGCAGCGCCTGCTGCTCGAGTTCGTCGTCGGCGAGGTCGCCGAAGTACTCGGCCACGC 13271

Qy         84 CACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCAC 143
              |  |    ||   |||  | |||| || |  |||| ||||||||||||||||||||| ||
Db      13272 CCGCGGTCACCGGATCGACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGAC 13331

Qy        144 CGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCT 203
              |||||||||||| |||||||||||  |  | ||   || |||  ||||||  | |||||
Db      13332 CGCCGTCGAACTCCGCAACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGCGACCCT 13391

Qy        204 CGCCTTCGACCACCCCAACCCCACCACCCTCACCCACCACCT 245
               | ||||||||||| | |||  |  | ||| ||  ||||||
Db      13392 GGTCTTCGACCACCCAAGCCCGGCGGCACTCGCCTCCCACCT 13433


RESULT 13
US-07-642-734C-3
; Sequence 3, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, L
;    APPLICANT:  Donadio, S
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Recombinant DNA Method for Producing
;    TITLE OF INVENTION:  Erythromycin Analogs
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Danckers, Andreas M
;      REGISTRATION NUMBER:  32652
;      REFERENCE/DOCKET NUMBER:  4952.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20235 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 238
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  19..10722
;      OTHER INFORMATION:  /codon_start= 19
;      OTHER INFORMATION:  /function= "gene eryA"
;      OTHER INFORMATION:  /product= "eryA ORF2 encoding modules 3 & 4 for
;      OTHER INFORMATION:  6-deoxyerythronolide B"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  19..4470
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  97..1482
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACP synthase of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1693..2670
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3406..3921
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4171..4428
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 3"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4471..10722
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 4"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  4471..5847
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACPsynhase domain of module"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6054..7026
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 4"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  7165..9216
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  dehydratase and enoylreductase domains m"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9433..9984
;      OTHER INFORMATION:  /function= "approximate span
;      OTHER INFORMATION:  beta-ketoreductase of module 4"
```

```
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-19

  Query Match             48.2%;  Score 118.6;  DB 3;  Length 38506;
  Best Local Similarity   67.8%;  Pred. No. 3e-15;
  Matches  166;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||||    |||||  ||  | | |  | |   ||| |  | ||| |||||  | |||
Db      24507 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 24566

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
               | || ||||||||||| |||  ||||||    |     |    ||  || || || ||||||
Db      24567 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 24626

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||  ||| ||||||||| |||||||||||||||||| |||||||| |||   |  | | |||
Db      24627 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 24686

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || || || || ||  |||| || | ||||||||||||||| |||        ||| ||
Db      24687 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 24746

Qy        241 CACCT 245
              |||||
Db      24747 CACCT 24751


RESULT 10
US-09-428-517-1
; Sequence 1, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Shah, Sanjay Krishnakant
;  APPLICANT: McDaniel, Robert
;  APPLICANT: Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/428,517
;  CURRENT FILING DATE: 1999-10-28
;  EARLIER APPLICATION NUMBER: 60/120,254
;  EARLIER FILING DATE: 1999-02-16
;  EARLIER APPLICATION NUMBER: 60/106,100
;  EARLIER FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 50937
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artifical Sequence: Recombinant DNA
US-09-428-517-1

  Query Match             48.0%;  Score 118.2;  DB 4;  Length 50937;
  Best Local Similarity   72.5%;  Pred. No. 3.6e-15;
  Matches  153;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy         35 CCCTCCTCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACACCACCCCCGACA 94
              |||| | || ||| ||||||  |||| |||||   || || || ||||    ||||||
Db      29173 CCCTGCGCGACCTCGTCCGCGGCCACGTCGCCCTGGTGCTCGGACACAGTGGCCCCGAGG 29232

Qy         95 CCATCCCCCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCACCGCCGTCGAAC 154
              |||||  | |||| |  || ||||| ||| |||| ||||||||||| ||||| |||||||
Db      29233 CCATCGACGCCGAACAGGCCTTCCGGGACATCGGTTTCGACTCCCTGACCGCAGTCGAAC 29292

Qy        155 TACGCAACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCTCGCCTTCGACC 214
              |  | ||||||||   |  |   ||||| |||||| | |||  ||| |||| ||||||
Db      29293 TCAGAAACCGGCTGAACGCCGAGACCGGCCTCCGCTTGCCCGGCACGCTCGTGTTCGACT 29352

Qy        215 ACCCCAACCCCACCACCCTCACCCACCACCT 245
              |||||||||| | | | ||| || | |||||
Db      29353 ACCCCAACCCGAGCGCGCTCGCCGATCACCT 29383


RESULT 11
US-09-105-537-34
; Sequence 34, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 34
;   LENGTH: 4689
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-34

  Query Match             45.2%;  Score 111.2;  DB 4;  Length 4689;
  Best Local Similarity   70.3%;  Pred. No. 7.9e-14;
  Matches  149;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy         35 CCCTCCTCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACACCACCCCCGACA 94
              | |||||| |||| |||||  |||||   ||  |||| || |||||  || ||||||||
Db       4211 CGCTCCTCACCCTCGTCCGTACCCACGCGGCGGCCGTACTCGGCCATTCCTCCCCCGACC 4270

Qy         95 CCATCCCCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCACCGCCGTCGAAC 154
                |  |||||| ||| || |||  ||| |||||||||||||| || || |||||  | |
Db       4271 GGGTGGCCCCCGGCCGTGCCTTCACCGAGCTCGGCTTCGACTCGCTGACGGCCGTGCAGC 4330

Qy        155 TACGCAACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCTCGCCTTCGACC 214
              | ||||||| |||||||         |||   | | |||||| |||||   | ||||||||
Db       4331 TCCGCAACCAGCTCTCCACGGTGGTCGGCAACAGGCTCCCCGCCACCACGGTCTTCGACC 4390

Qy        215 ACCCCAACCCCACCACCCTCACCCACCACCTC 246
              |||| |  ||| || | ||| ||   ||||||
Db       4391 ACCCGACGCCCGCCGCACTCGCCGCGCACCTC 4422
```

```
Db    2294792 TA 2294791

RESULT 6
US-09-434-288-7
; Sequence 7, Application US/09434288
; Patent No. 6303767
; GENERAL INFORMATION:
;  APPLICANT:  Betlach C., Melanie
;  APPLICANT:  McDaniel, Robert
;  TITLE OF INVENTION:  POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
;  TITLE OF INVENTION:  CONSTRUCTS THEREFOR
;  FILE REFERENCE:  30062-20030.00
;  CURRENT APPLICATION NUMBER: US/09/434,288
;  CURRENT FILING DATE:  1999-11-05
;  PRIOR APPLICATION NUMBER:  60/107,093
;  PRIOR FILING DATE:  1998-11-05
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 7
;   LENGTH: 1681
;   TYPE: DNA
;   ORGANISM: Streptomyces narbonensis
US-09-434-288-7

  Query Match             49.5%;  Score 121.8;  DB 4;  Length 1681;
  Best Local Similarity   68.6%;  Pred. No. 7.3e-16;
  Matches  168;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||||    ||| |  ||  |   |  | |   ||| |  | ||| |||||  | |||
Db        541 CTCGCCGCGATGACCCCGGACGACAGGACCGCGCACCTGCGGGACCTCGTCCGTACGCAC 600

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 120
                | || |||||||||| |||  | ||||    |    ||  ||  || ||||| ||||||
Db        601 GTGGCGACCGTCCTGGGACACGGCGCCCCGAGCCGGGTCGACCTGGAGCGCGCCTTCCGC 660

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||  ||| ||||||||||||||||||||||| || ||||||||| |||  |  | |||||
Db        661 GACACCGGTTTCGACTCCCTCACCGCCGTCGAGCTCCGCAACCGCCTCAACGCCGCCACC 720

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || || ||||||||  |||| |||| ||||||||||||||  ||        ||| ||
Db        721 GGGCTGCGCCTCCCGGCCACGCTCGTCTTCGACCACCCCACTCCGGGGGAGCTCGCCGGG 780

Qy        241 CACCT 245
              |||||
Db        781 CACCT 785


RESULT 7
US-09-105-537-32
; Sequence 32, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-32

  Query Match             48.2%;  Score 118.6;  DB 4;  Length 11220;
  Best Local Similarity   67.8%;  Pred. No. 3e-15;
  Matches  166;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||||    |||||  ||  | | |  | |   ||| |  | ||| |||||  | |||
Db      10678 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 10737

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 120
                | || |||||||||| |||  ||||||    |    ||  || || || ||||||
Db      10738 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 10797

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||  ||| |||||||| ||||||||||||||||| ||||||||  |||  |  | | |||
Db      10798 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 10857

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || || || || ||  |||| || | |||||||||||||| |||        ||| ||
Db      10858 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 10917

Qy        241 CACCT 245
              |||||
Db      10918 CACCT 10922


RESULT 8
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-5

  Query Match             48.2%;  Score 118.6;  DB 4;  Length 36778;
  Best Local Similarity   67.8%;  Pred. No. 3e-15;
  Matches  166;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||||    |||||  ||  | | |  | |   ||| |  | ||| |||||  | |||
Db      26365 CTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCAC 26424

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 120
                | || |||||||||| |||  ||||||    |    ||  || || || ||||||
Db      26425 GTGGCGACCGTCCTGGGACACGGCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTTCCGC 26484

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||  ||| |||||||| ||||||||||||||||| ||||||||  |||  |  | | |||
Db      26485 GACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGCGACC 26544

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || || || || ||  |||| || | |||||||||||||| |||        ||| ||
Db      26545 GGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGCCGGG 26604

Qy        241 CACCT 245
              |||||
Db      26605 CACCT 26609


RESULT 9
US-09-320-878-19
; Sequence 19, Application US/09320878A
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  43280 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  816..14234
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14351..19945
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20010..31199
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31232..36067
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36249..41774
US-08-804-227C-1

  Query Match            52.1%;  Score 128.2;  DB 2;  Length 43280;
  Best Local Similarity  70.2%;  Pred. No. 4.4e-17;
  Matches  172;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||| |||| |  ||    |  ||  |||| ||  | |||| |||  | ||||| |||||
Db      30648 CTGACCGGGCTCACGGCACAGGAAGGGCACGCCCTCGTCCTGGCCGAGATCCGCGCCCAC 30707

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
                ||  | ||  |||||||||  | ||  |||| | ||||||   ||||| || |||
Db      30708 GCGGCGGCGGTGCTGGGCCACGGCTCCGACGACTCGATCCCCGAGGACCGGGCCTTCAAG 30767

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||||||||||||||| ||||||||||| ||  | ||||||||||    |    |||||
Db      30768 GACCTCGGCTTCGACTCGCTCACCGCCGTGGAGATGCGCAACCGGCTGAGCGCGGCCACC 30827

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || ||||| || ||| |||||||| |||||||||| | |||    |   ||  ||  |
Db      30828 GGCCTCCGGCTGCCCGCCACCCTCGTCTTCGACCACCCGACCCCGGGCGAGCTGGCCGGC 30887

Qy        241 CACCT 245
              |||||
Db      30888 CACCT 30892


RESULT 4
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match            50.2%;  Score 123.6;  DB 4;  Length 4403765;
  Best Local Similarity  69.4%;  Pred. No. 3.4e-16;
  Matches  168;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||||  || | | |  | ||  | || ||| ||  ||| || | |||||| ||||| |||
Db    2297332 CTGGAAGGACTGCCCGAAGACGAGCAACACGCCGTCCTGCTGGACCTGGTGCGCTCGCAC 2297273

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
              |||||||||||| |||||   | ||| ||||||  |||||  ||| ||  | || ||||
Db    2297272 ATCGCCACCGTGCTGGGTAGCGCCAGCCCCGAAGCCATCGACCCGGATAGGGCATTCCAG 2297213

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||  | || |||||||| |||||||| |||||| | |||||||||||      | |||||
Db    2297212 GAGTTGGGTTTCGACTCGCTCACCGCGGTCGAAATGCGCAACCGGCTCAAATCCGCCACC 2297153

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || ||   ||  |  |||| |||  ||||||| |||||||| || |  | ||  ||  |
Db    2297152 GGCCTGGCGCTTTCACCCACGCTCATCTTCGACTACCCCAACTCCGCGGCGCTGGCCGGC 2297093

Qy        241 CA 242
              ||
Db    2297092 TA 2297091


RESULT 5
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match            50.2%;  Score 123.6;  DB 4;  Length 4411529;
  Best Local Similarity  69.4%;  Pred. No. 3.4e-16;
  Matches  168;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||||  || | | |  | ||  | || ||| ||  ||| || | |||||| ||||| |||
Db    2295032 CTGGAAGGACTGCCCGAAGACGAGCAACACGCCGTCCTGCTGGACCTGGTGCGCTCGCAC 2294973

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
              |||||||||||| |||||   | ||| ||||||  |||||  ||| ||  | || ||||
Db    2294972 ATCGCCACCGTGCTGGGTAGCGCCAGCCCCGAAGCCATCGACCCGGATAGGGCATTCCAG 2294913

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||  | || |||||||| |||||||| |||||| | |||||||||||      | |||||
Db    2294912 GAGTTGGGTTTCGACTCGCTCACCGCGGTCGAAATGCGCAACCGGCTCAAATCCGCCACC 2294853

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || ||   ||  |  |||| |||  ||||||| |||||||| || |  | ||  ||  |
Db    2294852 GGCCTGGCGCTTTCACCCACGCTCATCTTCGACTACCCCAACTCCGCGGCGCTGGCCGGC 2294793

Qy        241 CA 242
```

```
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-227C-7

  Query Match             55.8%;  Score 137.2;  DB 2;  Length 44377;
  Best Local Similarity   72.4%;  Pred. No. 8.4e-19;
  Matches  178;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||| |||  ||| |   | |  | | |||| ||||||| ||    ||||| ||| | |||
Db      24427 CTGACCGCGCAGTCCGGCCACAAGCGGCACGCCACCCTGCTGCGGCTGGTGCGCGCACAC 24486

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
              |||| ||||||| || ||  || ||  |||| |  |   |  || |||||| ||||||
Db      24487 GCCGCCGCCGTCCTCGGACAGTCCTCCGGCGACGCGGTGAGCAGCGCCCGCGCCTTCCGC 24546

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||||||||||||||| || |||||| ||||||| ||| ||||||||  |  || ||||
Db      24547 GACCTCGGCTTCGACTCGCTGACCGCCCTCGAACTGCGCGACCGGCTCAGCACCAGCACC 24606

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || |||   || |||||| |||| | |||||||||| ||| |||  || | ||| |||
Db      24607 GGGCTCAAACTGCCCACCTCCCTGGTCTTCGACCACTCCAGCCCGGCCGCGCTCGCCCGG 24666

Qy        241 CACCTC 246
              ||||||
Db      24667 CACCTC 24672


RESULT 2
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-198-1

  Query Match             55.8%;  Score 137.2;  DB 2;  Length 44377;
  Best Local Similarity   72.4%;  Pred. No. 8.4e-19;
  Matches  178;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||| |||  ||| |   | |  | | |||| ||||||| ||    ||||| ||| | |||
Db      24427 CTGACCGCGCAGTCCGGCCACAAGCGGCACGCCACCCTGCTGCGGCTGGTGCGCGCACAC 24486

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
              |||| ||||||| || ||  || ||  |||| |  |   |  || |||||| ||||||
Db      24487 GCCGCCGCCGTCCTCGGACAGTCCTCCGGCGACGCGGTGAGCAGCGCCCGCGCCTTCCGC 24546

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||||||||||||||| || |||||| ||||||| ||| ||||||||  |  || ||||
Db      24547 GACCTCGGCTTCGACTCGCTGACCGCCCTCGAACTGCGCGACCGGCTCAGCACCAGCACC 24606

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || |||   || |||||| |||| | |||||||||| ||| |||  || | ||| |||
Db      24607 GGGCTCAAACTGCCCACCTCCCTGGTCTTCGACCACTCCAGCCCGGCCGCGCTCGCCCGG 24666

Qy        241 CACCTC 246
              ||||||
Db      24667 CACCTC 24672


RESULT 3
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 10.6565 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_11413_11658
Perfect score: 246
Sequence: 1 ctggccggccagacacacga.....ccaccctcacccaccacctc 246

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	137.2	55.8	44377	2	US-08-804-227C-7	Sequence 7, Appli
	2	137.2	55.8	44377	2	US-08-804-198-1	Sequence 1, Appli
	3	128.2	52.1	43280	2	US-08-804-227C-1	Sequence 1, Appli
c	4	123.6	50.2	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	5	123.6	50.2	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	6	121.8	49.5	1681	4	US-09-434-288-7	Sequence 7, Appli
	7	118.6	48.2	11220	4	US-09-105-537-32	Sequence 32, Appl
	8	118.6	48.2	36778	4	US-09-105-537-5	Sequence 5, Appli
	9	118.6	48.2	38506	3	US-09-320-878-19	Sequence 19, Appl
	10	118.2	48.0	50937	4	US-09-428-517-1	Sequence 1, Appli
	11	111.2	45.2	4689	4	US-09-105-537-34	Sequence 34, Appl
	12	110	44.7	13842	4	US-09-105-537-30	Sequence 30, Appl
	13	108.8	44.2	20235	1	US-07-642-734C-3	Sequence 3, Appli
	14	108.8	44.2	20235	3	US-08-439-009A-3	Sequence 3, Appli
	15	107.4	43.7	80161	3	US-09-036-987A-1	Sequence 1, Appli
	16	107.4	43.7	80161	4	US-09-370-700-1	Sequence 1, Appli
c	17	105.2	42.8	1434	4	US-09-434-288-3	Sequence 3, Appli
	18	104.4	42.4	13987	2	US-08-804-227C-13	Sequence 13, Appl
	19	103.6	42.1	15872	4	US-09-105-537-1	Sequence 1, Appli
	20	102	41.5	49377	1	US-08-764-233A-1	Sequence 1, Appli
	21	100.4	40.8	11219	1	US-07-642-734C-1	Sequence 1, Appli
	22	100.4	40.8	11219	3	US-08-439-009A-1	Sequence 1, Appli
	23	99.2	40.3	28958	1	US-08-258-261B-6	Sequence 6, Appli
	24	99.2	40.3	28958	1	US-08-456-837-6	Sequence 6, Appli
	25	99.2	40.3	28958	1	US-08-457-342-6	Sequence 6, Appli
	26	99.2	40.3	28958	1	US-08-457-646A-6	Sequence 6, Appli
	27	99.2	40.3	28958	1	US-08-458-076A-6	Sequence 6, Appli
	28	99.2	40.3	28958	1	US-08-764-233A-4	Sequence 4, Appli
	29	99.2	40.3	28958	1	US-08-457-335A-6	Sequence 6, Appli
	30	99.2	40.3	28958	1	US-08-729-214-6	Sequence 6, Appli
	31	99.2	40.3	28958	3	US-09-028-934-6	Sequence 6, Appli
	32	95.6	38.9	4041	4	US-09-105-537-36	Sequence 36, Appl
	33	87.6	35.6	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	34	87.6	35.6	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	35	86.2	35.0	561	3	US-09-154-083-13	Sequence 13, Appl
c	36	78.6	32.0	748	3	US-09-154-083-1	Sequence 1, Appli
c	37	78.6	32.0	753	3	US-09-154-083-28	Sequence 28, Appl
c	38	77.6	31.5	318	4	US-09-165-264-12	Sequence 12, Appl
c	39	74.4	30.2	320	4	US-09-165-264-14	Sequence 14, Appl
c	40	73.8	30.0	319	4	US-09-165-264-8	Sequence 8, Appli
c	41	73.8	30.0	320	4	US-09-165-264-7	Sequence 7, Appli
c	42	73.6	29.9	320	4	US-09-165-264-13	Sequence 13, Appl
c	43	73.4	29.8	320	4	US-09-165-264-11	Sequence 11, Appl
	44	71.6	29.1	33529	4	US-09-144-085-3	Sequence 3, Appli
c	45	70.6	28.7	367	4	US-09-060-756-627	Sequence 627, App

ALIGNMENTS

RESULT 1
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rosteck, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCI(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:

PA (STRO/) STROM A R.
XX
PI Zotchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR WPI; 2001-557614/62.
DR P-PSDB; AAE10153.
XX
PT New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
PS Example 2; Page 186-187; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrolide antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a ERD48 insert DNA in pKR48 which is used in an
CC exemplification of the invention.
XX
SQ Sequence 2700 BP; 324 A; 1087 C; 876 G; 413 T; 0 other;

Query Match 50.2%; Score 123.4; DB 22; Length 2700;
Best Local Similarity 73.7%; Pred. No. 7.9e-13;
Matches 157; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Qy 33 CACCCTCCTCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACACCACCCCCGA 92
Db 2006 CACCCTCCTCGCGCTCGTCCGCGACCAGGCCGCGATGGTCCTCGGCCACACCTCGGGCGA 2065

Qy 93 CACCATCCCCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCACCGCCGTCGA 152
Db 2066 CGGCGTCGACCCGTCCCGCGCCTTCCGCGACCTCGGCTTCGACTCGCTCACCGCGGTCGA 2125

Qy 153 ACTACGCAACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCTCGCCTTCGA 212
Db 2126 ACTCCGCAACCGCATCGGCGCGGCCACCGGCCTGCGGCTACCGGCCACGGCCGTCTTCGA 2185

Qy 213 CCACCCCAACCCCACCACCCTCACCCACCACCT 245
Db 2186 CTACCCCACCGCCGATGCCCTCGCCGCACACCT 2218

RESULT 15
AAS18438
ID AAS18438 standard; DNA; 1681 BP.
XX
AC AAS18438;
XX
DT 12-MAR-2002 (first entry)
XX
DE Contig 115 DNA encoding S. narbonensis polyketide synthase.
XX
KW Narbonolide polyketide synthase; PKS; narbomycin modification enzyme;
KW erythromycin; rapamycin; tylosin; picromycin; methylmycin; neomethymycin;
KW agriculture; ds.
XX
OS Streptomyces narbonensis.
XX
PN US6303767-B1.
XX
PD 16-OCT-2001.
XX
PF 05-NOV-1999; 99US-0434288.
XX
PR 05-NOV-1998; 98US-107093P.
PR 27-MAY-1999; 99US-0320878.
XX
PA (KOSA-) KOSAN BIOSCIENCES INC.
XX
PI Betlach MC, McDaniel R;
XX
DR WPI; 2002-065495/09.
XX
PT Nucleic acids encoding narbonolide polyketide synthases from
PT Streptomyces narbonensis, useful for the recombinant production of
PT polyketides, e.g. narbomycin -
XX
PS Claim 1; Column 18; 24pp; English.
XX
CC The present invention relates to recombinant DNA vectors (cosmids)
CC that encode for the narbonolide polyketide synthase (PKS) enzyme and
CC various narbomycin modification enzymes from Streptomyces narbonensis.
CC The recombinant DNA vectors can be used to produce recombinant ketide
CC synthases and a variety of different polyketides (e.g. erythromycin,
CC rapamycin, tylosin, narbomycin, picromycin, methylmycin and
CC neomethymycin) for use in agriculture, medicine and health. The
CC recombinant vectors may be used to produce polyketides in relatively
CC high yields. AAS18432-AAS18443 represent contig DNA sequences that
CC encode for S. narbonensis PKS enzymes.
XX
SQ Sequence 1681 BP; 231 A; 672 C; 580 G; 198 T; 0 other;

Query Match 49.5%; Score 121.8; DB 24; Length 1681;
Best Local Similarity 68.6%; Pred. No. 1.6e-12;
Matches 168; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
Db 541 CTCGCCGCGATGACCCCGGACGACAGGACCGCGCACCTGCGGGACCTCGTCCGTACGCAC 600

Qy 61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
Db 601 GTGGCGACCGTCCTGGGACACGGCGCCCCGAGCCGGGTCGACCTGGAGCGCGCCTTCCGC 660

Qy 121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
Db 661 GACACCGGTTTCGACTCCCTCACCGCCGTCGAGCTCCGCAACCGCCTCAACGCCGCCACC 720

Qy 181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
Db 721 GGGCTGCGCCTCCCGGCCACGCTCGTCTTCGACCACCCCACTCCGGGGGAGCTCGCCGGG 780

Qy 241 CACCT 245
Db 781 CACCT 785

Search completed: June 17, 2003, 13:49:16
Job time : 61.0269 secs

```
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

  Query Match           50.2%;  Score 123.6;  DB 22;  Length 4403765;
  Best Local Similarity   69.4%;  Pred. No. 2.7e-13;
  Matches  168;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              |||| || | | |  | ||  | || ||| ||  ||| || | |||||| ||||| |||
Db    2297332 CTGGAAGGACTGCCCGAAGACGAGCAACACGCCGTCCTGCTGGACCTGGTGCGCTCGCAC 2297273

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
              |||||||||||| |||||   | ||| ||||||  |||||  ||| ||  | || ||||
Db    2297272 ATCGCCACCGTGCTGGGTAGCGCCAGCCCCGAAGCCATCGACCCGGATAGGGCATTCCAG 2297213

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||  | || |||||||| |||||||| |||||| | ||||||||||||     | |||||
Db    2297212 GAGTTGGGTTTCGACTCGCTCACCGCGGTCGAAATGCGCAACCGGCTCAAATCCGCCACC 2297153

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || ||    ||  |  |||| ||| ||||||| |||||||| || |  | ||  ||  |
Db    2297152 GGCCTGGCGCTTTCACCCACGCTCATCTTCGACTACCCCAACTCCGCGGCGCTGGCCGGC 2297093

Qy        241 CA 242
              |
Db    2297092 TA 2297091


RESULT 13
AAI99682/c
ID   AAI99682 standard; DNA; 4411529 BP.
XX
AC   AAI99682;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fleischmann RD,  White OR,  Fraser CM,  Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ  -
XX
PS   Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

  Query Match           50.2%;  Score 123.6;  DB 22;  Length 4411529;
  Best Local Similarity   69.4%;  Pred. No. 2.7e-13;
  Matches  168;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              |||| || | | |  | ||  | || ||| ||  ||| || | |||||| ||||| |||
Db    2295032 CTGGAAGGACTGCCCGAAGACGAGCAACACGCCGTCCTGCTGGACCTGGTGCGCTCGCAC 2294973

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
              |||||||||||| |||||   | ||| ||||||  |||||  ||| ||  | || ||||
Db    2294972 ATCGCCACCGTGCTGGGTAGCGCCAGCCCCGAAGCCATCGACCCGGATAGGGCATTCCAG 2294913

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||  | || |||||||| |||||||| |||||| | ||||||||||||     | |||||
Db    2294912 GAGTTGGGTTTCGACTCGCTCACCGCGGTCGAAATGCGCAACCGGCTCAAATCCGCCACC 2294853

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || ||    ||  |  |||| ||| ||||||| |||||||| || |  | ||  ||  |
Db    2294852 GGCCTGGCGCTTTCACCCACGCTCATCTTCGACTACCCCAACTCCGCGGCGCTGGCCGGC 2294793

Qy        241 CA 242
              |
Db    2294792 TA 2294791


RESULT 14
AAD17193
ID   AAD17193 standard; DNA; 2700 BP.
XX
AC   AAD17193;
XX
DT   29-NOV-2001  (first entry)
XX
DE   ERD48 insert DNA in pKR48.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..2699
FT                   /*tag=  a
FT                   /product= "Protein encoded by ERD48 insert DNA"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
```

```
Db      26912 GGCCTGCACCTGTCGGCCACCATGGTCTTCGACCACCCCACCCTGTCCGCCCTCGCGGAG 26971

Qy        241 CACCT 245
              |||||
Db      26972 CACCT 26976


RESULT 11
AAT80413
ID   AAT80413 standard; DNA; 43280 BP.
XX
AC   AAT80413;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Tylactone synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             816..14243
FT                   /*tag= a
FT                   /transl_except= (pos: 816..818, aa: Met)
FT                   /note= "ORF1 encodes protein shown in AAW22601"
FT   CDS             14351..19945
FT                   /*tag= b
FT                   /transl_except= (pos: 14351..14353, aa: Met)
FT                   /note= "ORF2 encodes protein shown in AAW22602"
FT   CDS             20010..31199
FT                   /*tag= c
FT                   /transl_except= (pos: 20010..20012, aa: Met)
FT                   /note= "ORF3 encodes protein shown in AAW22603"
FT   CDS             31232..36067
FT                   /*tag= d
FT                   /note= "ORF4 encodes protein shown in AAW22604"
FT   CDS             36249..41774
FT                   /*tag= e
FT                   /note= "ORF5 encodes protein shown in AAW22605"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22601-W22605.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Claim 2; Pages 8-66; 220pp; English.
XX
CC   This sequence represents the tylactone synthase gene cluster of the
CC   invention. This sequence is also referred to as the tylG gene, and was
CC   isolated from Streptomyces fradiae. This sequence encodes multifunctional
CC   proteins which direct the synthesis of the polyketide tylactone, isolated
CC   from Streptomyces fradiae. Tylactone is the basic building block of the
CC   antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC   type of carboxylic acids incorporated, the number of carboxylic acids
CC   incorporated and/or the post-condensation reactions performed, thereby
CC   resulting in novel tylosin-related polyketides.
XX
SQ   Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;

  Query Match           52.1%;  Score 128.2;  DB 18;  Length 43280;
  Best Local Similarity  70.2%;  Pred. No. 8.6e-14;
  Matches  172;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||| |||| |  ||      |  ||  |||| ||  | |||| |||  | ||||| |||||
Db      30648 CTGACCGGGCTCACGGCACAGGAAGGGCACGCCCTCGTCCTGGCCGAGATCCGCGCCCAC 30707

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 120
                ||  | ||  |||||||||  | ||  |||| | ||||||   ||||| || |||
Db      30708 GCGGCGGCGGTGCTGGGCCACGGCTCCGACGACTCGATCCCCGAGGACCGGGCCTTCAAG 30767

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||||||||||||||| |||||||||| ||  | ||||||||||   |    |||||
Db      30768 GACCTCGGCTTCGACTCGCTCACCGCCGTGGAGATGCGCAACCGGCTGAGCGCGGCCACC 30827

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || ||||| || ||| |||||||| |||||||||||| | |||    |   ||  || |
Db      30828 GGCCTCCGGCTGCCCGCCACCCTCGTCTTCGACCACCCGACCCCGGGCGAGCTGGCCGGC 30887

Qy        241 CACCT 245
              |||||
Db      30888 CACCT 30892


RESULT 12
AAI99683/c
ID   AAI99683 standard; DNA; 4403765 BP.
XX
AC   AAI99683;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fleischmann RD,  White OR,  Fraser CM,  Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ  -
XX
PS   Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
```

```
  Query Match             54.1%;  Score 133;  DB 22;  Length 125401;
  Best Local Similarity   71.4%;  Pred. No. 1.2e-14;
  Matches  175;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||||  ||| |  ||   | ||  || ||| ||| || || ||||||  |    |
Db      29738 CTCGCCGAGCAGCCCGCCGCGGACCACGACGCCATCCTGCTGGCACTGGTCACCGAGAAG 29797

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
                ||| |||||||| ||||||  ||  || ||||    ||   |||||||   | ||||||
Db      29798 GCCGCGACCGTCCTCGGCCACGCCGACCACGACCTGCTCGAACCCGACCTGCCCTTCCGC 29857

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||||| |||||||||||| || ||||| ||||| || ||||||| |||| ||      |||
Db      29858 GACCTGGGCTTCGACTCGCTGACCGCGGTCGACCTGCGCAACCAGCTCACCGCGGAAACC 29917

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || |||  ||| ||| ||||||||| |||||||||||||||||  || ||||| ||  |
Db      29918 GGCCTCACCCTGCCCGCCACCCTCGTCTTCGACCACCCCAACCCGGCCGCCCTCGCCGCC 29977

Qy        241 CACCT 245
              |||||
Db      29978 CACCT 29982


RESULT 10
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..1035)
FT                   /*tag=  a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement (1056..2576)
FT                   /*tag=  b
FT                   /product= "NysD1 protein"
FT   CDS             2806..6906
FT                   /*tag=  c
FT                   /product= "NysA protein"
FT   CDS             6952..16530
FT                   /*tag=  d
FT                   /product= "NysB protein"
FT   CDS             16550..49840
FT                   /*tag=  e
FT                   /product= "NysC protein"
FT   CDS             50260..51015
FT                   /*tag=  f
FT                   /product= "NysE protein"
FT   CDS             51405..54305
FT                   /*tag=  g
FT                   /product= "NysR1 protein"
FT   CDS             54329..57190
FT                   /*tag=  h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180..59963
FT                   /*tag=  i
FT                   /product= "NysR3 protein"
FT   CDS             60415..61047
FT                   /*tag=  j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736..62497
FT                   /*tag=  k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (62551..63615)
FT                   /*tag=  l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765..64961
FT                   /*tag=  m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 116-151; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

  Query Match             52.8%;  Score 129.8;  DB 22;  Length 65140;
  Best Local Similarity   70.6%;  Pred. No. 4.4e-14;
  Matches  173;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || | | | |     ||||  ||||  |||    | || |||| | ||||||||  |  |
Db      26732 CTCGGCCGACTCGACCACGCCCAACGCCACGAGGCACTGCTCGACATGGTCCGCAGCAGC 26791

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
                |||    |||||| |||||| ||   |  | || |||||  | ||||  |||||  ||||||
Db      26792 GCCGCGCTCGTCCTCGGCCACGCCGACGGCAACGCCATCGACGCCGAACGCGCCTTCCGC 26851

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||||||||||||||||| |||||||| |||||||| |||||||| ||    |  | |||||
Db      26852 GACCTCGGCTTCGACTCGCTCACCGCGGTCGAACTCCGCAACCGTCTGCGCACCGCCACC 26911

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || || | |||  |  ||||| | | |||||||||||||| ||   || ||||| |  |
```

```
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22606-W22610.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Example 2; Pages 110-134; 220pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. This sequence was used along with the tylG gene
CC   (see AAT80413) to create a hybrid ORF1 sequence (see AAT80415). The tylG
CC   gene is the tylactone synthase gene cluster of the invention. The tylG
CC   sequence was isolated from Streptomyces fradiae, and encodes
CC   multifunctional proteins which direct the synthesis of the polyketide
CC   tylactone. Tylactone is the basic building block of the antibiotic
CC   tylosin. The hybrid sequence can be used to transform S. ambofaciens
CC   lacking the srmG ORF1 sequence, or S. fradiae lacking the tylG ORF1
CC   sequence, so that they can produce polyketides. The DNA sequence can be
CC   modified so as to alter the type of carboxylic acids incorporated, the
CC   number of carboxylic acids incorporated and/or the post-condensation
CC   reactions performed, thereby resulting in novel tylosin-related
CC   polyketides.
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

  Query Match             55.8%;  Score 137.2;  DB 18;  Length 44377;
  Best Local Similarity   72.4%;  Pred. No. 2.6e-15;
  Matches  178;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||| |||  ||| |   | |  | | |||| |||||||  ||    |||||  ||| | |||
Db      24427 CTGACCGCGCAGTCCGGCCACAAGCGGCACGCCACCCTGCTGCGGCTGGTGCGCGCACAC 24486

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
                ||||  |||||||  ||  ||  ||  ||  ||||  |  |    |  ||  ||||||  ||||||
Db      24487 GCCGCCGCCGTCCTCGGACAGTCCTCCGGCGACGCGGTGAGCAGCGCCCGCGCCTTCCGC 24546

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||||||||||||||||  ||  ||||||  |||||||  |||  ||||||||  |  ||.||||
Db      24547 GACCTCGGCTTCGACTCGCTGACCGCCCTCGAACTGCGCGACCGGCTCAGCACCAGCACC 24606

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              ||  |||   ||  ||||||  ||||  |  ||||||||||  |||  |||   ||  |  |||  |||
Db      24607 GGGCTCAAACTGCCCACCTCCCTGGTCTTCGACCACTCCAGCCCGGCCGCGCTCGCCCGG 24666

Qy        241 CACCTC 246
              ||||||
Db      24667 CACCTC 24672


RESULT 9
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag= a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag= b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag= c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag= d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag= e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag= f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag= g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag= h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;
```

```
Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12230 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 12289

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12290 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 12349

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGA 152
              ||||||||||||||||||||||||||||||||
Db      12350 GACCTCGGCTTCGACTCCCTCACCGCCGTCGA 12381
```

```
RESULT 7
AAT78508
ID   AAT78508 standard; DNA; 44377 BP.
XX
AC   AAT78508;
XX
DT   26-FEB-1998  (first entry)
XX
DE   Platenolide synthase gene cluster.
XX
KW   Platenolide synthase gene cluster; platenolide production; srmG gene;
KW   multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   CDS             350..14002
FT                   /*tag=  a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in AAW23716"
FT   CDS             14046..20036
FT                   /*tag=  b
FT                   /note= "ORF2 encodes protein shown in AAW23717"
FT   CDS             20110..31284
FT                   /*tag=  c
FT                   /transl_except= (pos:20111..20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in AAW23718"
FT   CDS             31329..36071
FT                   /*tag=  d
FT                   /note= "ORF4 encodes protein shown in AAW23719"
FT   CDS             36155..41830
FT                   /*tag=  e
FT                   /note= "ORF5 encodes protein shown in AAW23720"
XX
PN   EP791656-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301066.
XX
PR   22-FEB-1996;   96US-0012050.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Burgett SG,  Kuhstoss SA,  Rao RN,  Richardson MA;
PI   Rosteck PR;
XX
DR   WPI; 1997-418047/39.
DR   P-PSDB; AAW23716-W23720.
XX
PT   DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT   for production of spiramycin-related polyketide antibiotics
XX
PS   Claim 9; Pages 8-33; 81pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. The DNA can be used to produce compounds
CC   exhibiting antibiotic activity based on the platenolide structure,
CC   including specifically the macrolide antibiotic spiramycin and spiramycin
CC   analogues and derivatives. Modifications of the platenolide synthase DNA
CC   sequence can be made so as to change the number and type of carboxylic
CC   acids incorporated into the growing polyketide chain and to change the
CC   kind of post-condensation processing that is conducted.
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;
```

```
  Query Match             55.8%;  Score 137.2;  DB 18;  Length 44377;
  Best Local Similarity   72.4%;  Pred. No. 2.6e-15;
  Matches  178;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||| |||  ||| |   | |  | | |||| ||||||| ||    ||||| ||| | |||
Db      24427 CTGACCGCGCAGTCCGGCCACAAGCGGCACGCCACCCTGCTGCGGCTGGTGCGCGCACAC 24486

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
                |||| ||||||| || ||  || ||  |||| |  |   |  || |||||| ||||||
Db      24487 GCCGCCGCCGTCCTCGGACAGTCCTCCGGCGACGCGGTGAGCAGCGCCCGCGCCTTCCGC 24546

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||||||||||||||| || |||||| ||||||| ||| |||||||| |  || ||||
Db      24547 GACCTCGGCTTCGACTCGCTGACCGCCCTCGAACTGCGCGACCGGCTCAGCACCAGCACC 24606

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || |||   || |||||| |||| | |||||||||| ||| |||  || | ||| |||
Db      24607 GGGCTCAAACTGCCCACCTCCCTGGTCTTCGACCACTCCAGCCCGGCCGCGCTCGCCCGG 24666

Qy        241 CACCTC 246
              ||||||
Db      24667 CACCTC 24672
```

```
RESULT 8
AAT80414
ID   AAT80414 standard; DNA; 44377 BP.
XX
AC   AAT80414;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Platenolide synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   platenolide synthase gene cluster; platenolide production; srmG gene;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   CDS             350..14002
FT                   /*tag=  a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in AAW22606"
FT   CDS             14046..20036
FT                   /*tag=  b
FT                   /note= "ORF2 encodes protein shown in AAW22607"
FT   CDS             20110..31284
FT                   /*tag=  c
FT                   /transl_except= (pos:20111..20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in AAW22608"
FT   CDS             31329..36071
FT                   /*tag=  d
FT                   /note= "ORF4 encodes protein shown in AAW22609"
FT   CDS             36155..41830
FT                   /*tag=  e
FT                   /note= "ORF5 encodes protein shown in AAW22610"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
```

```
Db      20269 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 20328

Qy        241 CACCTC 246
              ||||||
Db      20329 CACCTC 20334


RESULT 5
AAH79278
ID   AAH79278 standard; DNA; 31422 BP.
XX
AC   AAH79278;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 2.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag= a
FT                   /product= "AAG65266"
FT   CDS             14824..31422
FT                   /*tag= b
FT                   /product= "AAG65267"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65266, AAG65267.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Disclosure; Page 103-149; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;

  Query Match             99.3%;  Score 244.4;  DB 22;  Length 31422;
  Best Local Similarity   99.6%;  Pred. No. 2.8e-33;
  Matches  245;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20089 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 20148

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
              ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      20149 ATCGCCACCGTCCTCGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 20208

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20209 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 20268

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20269 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 20328

Qy        241 CACCTC 246
              ||||||
Db      20329 CACCTC 20334


RESULT 6
AAZ58381
ID   AAZ58381 standard; DNA; 12381 BP.
XX
AC   AAZ58381;
XX
DT   23-MAY-2000  (first entry)
XX
DE   Streptomyces avermitilis avermectin polyketide synthase modules 1+2.
XX
KW   Polyketide synthase; avermectin; insecticide; ss.
XX
OS   Streptomyces avermitilis.
XX
PN   WO200001827-A2.
XX
PD   13-JAN-2000.
XX
PF   06-JUL-1999;   99WO-GB02158.
XX
PR   06-JUL-1998;   98GB-0014622.
XX
PA   (BIOT-) BIOTICA TECHNOLOGY LTD.
PA   (PFIZ ) PFIZER INC.
XX
PI   Kellenberger JL,  Leadlay PF,  Staunton J,  Stutzman-Engwall KJ;
PI   McArthur HAI;
XX
DR   WPI; 2000-182117/16.
XX
PT   Mutated Type I polyketide synthase containing a polylinker site in an
PT   extension module for replacement of a reductive loop sequence, for
PT   producing polyketides, e.g. B1 avermectin -
XX
PS   Disclosure; Fig 7a-f; 75pp; English.
XX
CC   The present sequence is that of DNA encoding the first 2 modules
CC   of the avermectin polyketide synthase (PKS) of Streptomyces
CC   avermitilis.  The invention relates to nucleic acids encoding a
CC   Type I PKS such as avermectin in which a polylinker with multiple
CC   restriction sites replaces or 1 more PKS genes encoding enzymes
CC   associated with reduction.  Novel PKS are provided in which in
CC   which the reductive loop in a selected module of the Type I PKS is
CC   replaced with the equivalent segment from the same or different
CC   PKS gene cluster or by a mutated or synthetic segment.  Vectors and
CC   host cells, and methods for producing novel polyketides by
CC   culturing host cells are claimed.  The polyketides obtained are
CC   useful as antibiotics and insecticides.  Fermentation products
CC   containing C22-C23 dihydroavermectin, ivermectin and B1
CC   avermectins are claimed.
XX
SQ   Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;

  Query Match             61.8%;  Score 152;  DB 21;  Length 12381;
  Best Local Similarity   100.0%;  Pred. No. 1e-17;
  Matches  152;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag=  a
FT                   /product= "AAG65264"
FT                   /partial
FT   CDS             11971..30690
FT                   /*tag=  b
FT                   /product= "AAG65265"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65264, AAG65265.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Example 2; Page 58-123; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;

  Query Match             100.0%;  Score 246;  DB 22;  Length 30690;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-33;
  Matches  246;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11413 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 11472

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGCGTTCCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11473 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGCGTTCCGC 11532

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11533 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 11592

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11593 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 11652

Qy        241 CACCTC 246
              ||||||
Db      11653 CACCTC 11658


RESULT 4
AAA92302
ID   AAA92302 standard; DNA; 31422 BP.
XX
AC   AAA92302;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag=  a
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             14824..31422
FT                   /*tag=  b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23751, AAB23752.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 2; Page 134-203; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;

  Query Match             99.3%;  Score 244.4;  DB 21;  Length 31422;
  Best Local Similarity   99.6%;  Pred. No. 2.8e-33;
  Matches  245;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20089 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 20148

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGCGTTCCGC 120
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      20149 ATCGCCACCGTCCTCGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGCGTTCCGC 20208

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20209 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 20268

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65268.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Claim 10; Page 149-167; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is modified version of a
CC   fragment of the S. avermitilis genome.
XX
SQ   Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;

  Query Match             100.0%;  Score 246;  DB 22;  Length 11916;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-33;
  Matches  246;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11413 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 11472

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11473 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 11532

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11533 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 11592

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11593 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 11652

Qy        241 CACCTC 246
              ||||||
Db      11653 CACCTC 11658


RESULT 2
AAA92301
ID   AAA92301 standard; DNA; 30690 BP.
XX
AC   AAA92301;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag= a
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             11971..30690
FT                   /*tag= b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23749, AAB23750.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 2; Page 66-134; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;

  Query Match             100.0%;  Score 246;  DB 21;  Length 30690;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-33;
  Matches  246;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11413 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 11472

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11473 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 11532

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11533 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 11592

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11593 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 11652

Qy        241 CACCTC 246
              ||||||
Db      11653 CACCTC 11658


RESULT 3
AAH79277
ID   AAH79277 standard; DNA; 30690 BP.
XX
AC   AAH79277;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 1.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 48.936 Seconds
(without alignments)
11320.736 Million cell updates/sec

Title: US-09-914-286-1_COPY_11413_11658
Perfect score: 246
Sequence: 1 ctggccggccagacacacga.....ccaccctcacccaccacctc 246

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	246	100.0	11916	22	AAH79279	Streptomyces averm
	2	246	100.0	30690	21	AAA92301	S. avermitilis ave
	3	246	100.0	30690	22	AAH79277	Streptomyces averm
	4	244.4	99.3	31422	21	AAA92302	S. avermitilis ave
	5	244.4	99.3	31422	22	AAH79278	Streptomyces averm
	6	152	61.8	12381	21	AAZ58381	Streptomyces averm
	7	137.2	55.8	44377	18	AAT78508	Platenolide syntha
	8	137.2	55.8	44377	18	AAT80414	Platenolide syntha
	9	133	54.1	125401	22	AAD17186	Streptomyces nours
	10	129.8	52.8	65140	22	AAD17184	Streptomyces nours
	11	128.2	52.1	43280	18	AAT80413	Tylactone synthase
c	12	123.6	50.2	4403765	22	AAI99683	Mycobacterium tube
c	13	123.6	50.2	4411529	22	AAI99682	Mycobacterium tube
	14	123.4	50.2	2700	22	AAD17193	ERD48 insert DNA i
	15	121.8	49.5	1681	24	AAS18438	Contig 115 DNA enc
	16	118.6	48.2	11220	21	AAZ87298	S. venezuelae macr
	17	118.6	48.2	36778	21	AAZ87318	S. venezuelae pik
	18	118.6	48.2	37948	21	AAZ87285	S. venezuelae pik
	19	118.6	48.2	38506	21	AAA75633	Nucleotide sequenc
	20	118.6	48.2	38506	21	AAZ56001	Recombinant cosmid
	21	118.2	48.0	50937	21	AAA09469	Streptococcus olea
	22	118	48.0	27541	22	AAD17185	Streptomyces nours
	23	117	47.6	20394	22	AAF24892	Pimaricin biosynth
	24	112.2	45.6	53789	19	AAV21187	Amycolatopsis medi
	25	111.2	45.2	4689	21	AAZ87299	S. venezuelae macr
	26	110	44.7	13842	21	AAZ87297	S. venezuelae macr
	27	108.8	44.2	29879	14	AAQ46806	eryA region of S.
c	28	107.6	43.7	77536	21	AAA14651	Nucleotide sequenc
	29	107.4	43.7	9513	22	AAF88337	S. spinosa DNA fra
	30	107.4	43.7	47981	22	AAF30757	Micromonospora meg
	31	107.4	43.7	50000	22	AAF88313	S. spinosa DNA fra
	32	107.4	43.7	50000	22	AAF88316	S. spinosa DNA fra
	33	107.4	43.7	80161	20	AAZ21501	DNA fragment of Sa
c	34	105.2	42.8	1434	24	AAS18434	Contig 132 DNA enc
	35	104.4	42.4	13987	18	AAT80415	Hybrid srmG/tylG O
	36	103.6	42.1	15872	21	AAZ87283	S. venezuelae vep
	37	102	41.5	3978	21	AAC55785	Type I polyketide
	38	102	41.5	12249	21	AAC55840	Complete Mitomycin
	39	102	41.5	18331	21	AAC55857	Complete nucleotid
	40	102	41.5	49377	19	AAV05287	The soraphen biosy
	41	99.2	40.3	28598	17	AAT06769	Sorangium cellulos
	42	99.2	40.3	28958	18	AAT89956	Sorangium cellulos
	43	99.2	40.3	28958	21	AAA75299	DNA sequence of So
	44	98.8	40.2	15872	18	AAT68715	Streptomyces venez
	45	97.8	39.8	5676	19	AAV21186	Amycolatopsis medi

ALIGNMENTS

RESULT 1
AAH79279
ID AAH79279 standard; DNA; 11916 BP.
XX
AC AAH79279;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis coding sequence derivative SEQ ID NO: 3.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..11916
FT /*tag= a
FT /product= "AAG65268"
FT /partial
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX

/protein_id="AAK46387.1"
/db_xref="GI:13881781"
/translation="MVDQLQHATEALRKALVQVERLKRTNRALLERSSEPIAIVGMSC
RFPGGVDSPEGLWQMVADARDVMSEFPTDRGWDLAGLFDPDPDVRHKSYARTGGFVDG
VADFDPAFFGISPSEALAMDPQHRMLLELSWEALERAGIDPTGLRGSATGVFAGLIVG
GYGMLAEEIEGYRLTGMTSSVASGRVAYVLGLEGPAVSVDTACSSSLVALHMAVGSLR
SGECDLALAGGVTVNATPTVFVEFSRHRGLAPDGRCKPYAGRADGVGWSEGGGMLVLQ
RLSDARRLGHPVLAVVVGSAVNQDGASNGLTAPNGPSQQRVVRAALANAGLSAAEVDV
VEGHGTGTTLGDPIEAQALLATYGQDRGEPGEPLWLGSVKSNMGHTQAAAGVAGVIKM
VLAMRHELLPATLHVDVPSPHVDWSAGAVELLTAPRVWPAGARTRRAGVSSFGISGTN
AHVIIEAVPVVPRREAGWAGPVVPWVVSAKSESALRGQAARLAAYVRGDDGLDVADVG
WSLAGRSVFEHRAVVVGGDRDRLLAGLDELAGDQLGGSVVRGTATAAGKTVFVFPGQG
SQWLGMGIELLDTAPAFAQQIDACAEAFAEFVDWSLVDVLRGAPGAPGLDRVDVVQPV
LFAVMVSLAELWKSVAVHPDAVIGHSQGEIAAAYVAGALSLRDAARVVTLRSKLLAGL
AGPGGMVSIACGADQARDLLAPFGDRVSIAVVNGPSAVVVSGEVGALEELIAVCSTKE
LRTRRIEVDYASHSVEVEAIRGPLAEALSGIEPRSTRTVFFSTVTGNRLDTAGLDADY
WYRNVRQTVLFDQAVRNACEQGYRTFIESSPHPALITGVEETFAACTDGDSEAIVVPT
LGRGDGGLHRFLLSAASAFVAGVAVNWRGTLDGAGYVELPTYAFDKRRFWLSAEGSGA
DVSGLGLGASEHPLLGAVVDLPASGGVVLTGRLSPNVQPWLADHAVSDVVLFPGTGFV
ELAIRAGDEVGCSVLDELTLAAPLLLPATGSVAVQVVVDAGRDSNSRGVSIFSRADAQ
AGWLLHAEGILRPGSVEPGADLSVWPPAGAVTVDVADGYERLATRGYRYGPAFRGLTA
MWARGEEIFAEVRLPEAAGGVGGFGVHPALLDAVLHAVVIAGDPDELALPFAWQGVSL
HATGASAVRARIAPAGPSAVSVELADGLGLPVLSVASMVARPVTERQLLAAVSGSGPD
RLFEVIWSPASAATSPGPTPAYQIFESVAADQDPVAGSYVRSHQALAAVQSWLTDHES
GVLVVATRGAMALPREDVADLAGAAVWGLVRSAQTEHPGRIVLVDSDAATDDAAIAMA
LATGEPQVVLRGGQVYTARVRGSRAADAILVPPGDGPWRLGLGSAGTFENLRLEPVPN
ADAPLGPGQVRVAMRAIAANFRDIMITLGMFTHDALLGGEGAGVVVEVGPGVTEFSVG
DSVFGFFPDGSGTLVAGDVRLLLPMPADWSYAEAAAISAVFTTAYYAFIHLADVQPGQ
RVLIHAGTGGVGMAAVQLARHLGLEVFATASKGKWDTLRAMGFDDDHISDSRSLEFED
KFRAATGGRGFDVVLDSLAGEFVDASLRLVAPGGVFLEMGKTDIRDPGVIAQQYPGVR
YRAFDLFEPGRPRMHQYMLELATLFGDGVLRPLPVTTFDVRRAPAALRYLSQARHTGK
VVMLMPGSWAAGTVLITGGTGMAGSAVARHVVARHGVRNLVLVSRRGPDAPGAAELVA
ELAAAGAQVQVVACDAADRAALAKVIADIPVQHPLSGVIHTAGALDDAVVMSLTPDRV
DVVLRSKVDAAWHLHELTRDLDVSAFVMFSSMAGLVGSSGQANYAAANSFLDALAAHR
RAHGLPAISLGWGLWDQASAMTGGLDAADLARLGREGVLALSTAEALELFDTAMIVDE
PFLAPARIDLTALRAHAVAVPPMFSDLASAPTRRQVDDSVAAAKSKSALAHRLHGLPE
AEQHAVLLGLVRLHIATVLGNITPEAIDPDKAFQDLGFDSLTAVEMRNRLKSATGLSL
SPTLIFDYPTPNRLASYIRTELAGLPQEIKHTPAVRTTSEDPIAIVGMACRYPGGVNS
PDDMWDMLIQGRDVLSEFPADRGWDLAGLYNPDPDAAGACYTRTGGFVDGVGDFDPAF
FGVGPSEALAMDPQQRMLLELSWEALERAGIDPTGLRGSATGVFAGVMTQGYGMFAAE
PVEGFRLTGQLSSVASGRVAYVLGLEGPAVSVDTACSSSLVALHMAVGSLRSGECDLA
LAGGVTVNATPTVFVEFSRHRGLAPDGRCKPYAGRADGVGWSEGGGMLVLQRLSDARR
LGHPVLAVVVGSAVNQDGASNGLTAPNGPSQQRVVRAALANAGLSAAEVDVVEGHGTG
TTLGDPIEAQALLATYGQDRGEPGEPLWLGSVKSNMGHTQAAAGVAGVIKMVLAMRHE
LLPATLHVDVPSPHVDWSAGAVELLTAPRVWPAGARTRRAGVSSFGISGTNAHVIIEA
VPVVPRREAGWAGPVVPWVVSAKSESALRGQAARLAAYVRGDDGLDVADVGWSLAGRS
VFEHRAVVVGGDRDRLLAGLDELAGDQLGGSVVRGTATAAGKTVFVFPGQGSQWLGMG
MGLHAGYPVFAEAFNTVVGELDRHLLRPLREVMWGHDENLLNSTEFAQPALFAVEVAL
FRLLGSWGVRPDFVMGHSIGELSAAHVAGVLSLENAAVLVAARGRLMQALPAGGAMVA
VQAAEEEVRPLLSAEVDIAAVNGPASLVISGAQNAVAAVADQLRADGRRVHQLAVSHA
FHSPLMDPMIDEFAAVAAGIAIGRPTIGVISNVTGQLAGDDFGSAAYWRRHIRQAVRF
ADSVRFAQAAGGSRPLEVGPSGGLVASIEESLPDVAVTTMSALRKDRPEPATLTNAVA
QGFVTGMDLDWRAVVGEAQFVELPTYAFQRRRFWLSGDGVAADAAGLGLAASEHALLG
AVIDLPASGGVVLTGRLSPSVQGWLADHSVAGVTIFPGAGFVELAIRAGDEVGCGVVD
ELTLAAPLVLPASGSVAVQVVVNGPDESGVRGVSVYSRGDVGTGWVLHAEGALRAGSA
EPTADLAMWPPAGAVPVEVADGYQQLAERGYGYGPAFRGLTAMWRRGDEVFAEVALPA
DAGVSVTGFGVHPVLLDAALHAVVLSAESAERGQGSVLVPFSWQGVSLHAAGASAVRA
RIAPVGPSAVSIELADGLGLPVLSVASMLARPVTDQQLRAAVSSSGPDRLFEVTWSPQ
PSAAVEPLPVCAWGTTEDSAAVVFESVPLAGDVVAGVYAATSSVLDVLQSWLTRDGAG
VLVVMTRGAVALPGEDVTDLAGAAVWGLVRSAQTEHPGRIVLVDSDAPLDDSALAAVV
TTGEPQVLWRRGEVYTARVHGSRAVGGLLVPPSDRPWRLAMSTAGTFENLRLELIPDA
DAPLGPGQVRVAVSAIAANFRDVMIALGLYPDPDAVMGVEACGVVIETSLNKGSFAVG
DRVMGLFPEGTGTVASTDQRLLVKVPAGWSHTAAATTSVVFATAHYALVDLAAARSGQ
RVLIHAGTGGVGMAAVQLARHLGLEVFATASKGKWDTLRAMGFDDDHISDSRSLEFED
KFRAATGGRGFDVVLDSLAGEFVDASLRLVAPGGVFLEMGKTDIRDPGVIAQQYPGVR
YRAFDLFEAGPDRIAQILAELATLFGDGVLRPLPVTTFDVRCAPAALRYLSQARHTGK
VVMLMPGSWAAGTVLITGGTGMAGSAVARHVVARHGVRNLVLVSRRGPDAPGAAELVA
ELAAAGAQVQVVACDAADRAALAKVIADIPVQHPLSGVIHTAGALDDAVVMSLTPDRV
DVVLRSKVDAAWHLHELTRDLDVSAFVMFSSMAGLVGSSGQANYAAANSFLDALAAHR
RAHGLPAISLGWGLWDQASAMTGGLATVDFKRFARDGIVAMSSADALQLFDTAMIVDE
PFMLPAHIDFAALKVKFDGGTLPPMFVDLINAPTRRQVDDSLAAAKSKSALLQRLEGL
PEDEQHAVLLDLVRSHIATVLGSASPEAIDPDRAFQELGFDSLTAVEMRNRLKSATGL
ALSPTLIFDYPNSAALAGYMRRELLGSSPQDTSAVAAGEAELQRIVASIPVKRLRQAG

VLDLLLALANETETSGQDPALAPTAEQEIADMDLDDLVNAAFRNDDE"
BASE COUNT 3023 a 6045 c 4734 g 2168 t
ORIGIN

Query Match 50.2%; Score 123.6; DB 1; Length 15970;
Best Local Similarity 69.4%; Pred. No. 3.7e-07;
Matches 168; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```
Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||||  || | | |  | ||  | || ||| ||  ||| || | |||||| ||||| |||
Db       3835 CTGGAAGGACTGCCCGAAGACGAGCAACACGCCGTCCTGCTGGACCTGGTGCGCTCGCAC 3776

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
              |||||||||||| |||||  | ||| ||||||  |||||  ||| ||  | || ||||
Db       3775 ATCGCCACCGTGCTGGGTAGCGCCAGCCCCGAAGCCATCGACCCGGATAGGGCATTCCAG 3716

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||  | || ||||||||| |||||||| ||||||  | ||||||||||||      | |||||
Db       3715 GAGTTGGGTTTCGACTCGCTCACCGCGGTCGAAATGCGCAACCGGCTCAAATCCGCCACC 3656

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || ||    ||  |  |||| ||| ||||||| |||||||| || |  | ||  ||  |
Db       3655 GGCCTGGCGCTTTCACCCACGCTCATCTTCGACTACCCCAACTCCGCGGCGCTGGCCGGC 3596

Qy        241 CA 242
               |
Db       3595 TA 3594
```

Search completed: June 17, 2003, 17:55:53
Job time : 479.645 secs

```
                     /codon_start=1
                     /transl_table=11
                     /product="tylactone synthase module 3"
                     /protein_id="AAB66505.1"
                     /db_xref="GI:2317861"
                     /translation="MTAELVATRKRLGALEERAREPIAVVAMSCRYPGGVTTPEDLWR
                     LLADERDAVSGLPRDRGWDLDALYDPDGGPGTSYAREGGFLSHCAGFDAEFFGISPRE
                     ALAMDPQQRLLLETSWEALERAGVTADRARGSRTGVYAGVMYDDYGARVLYGAGAGPP
                     EDLEGYLVNGSAGSIASGRVSYTFGLRGPAVTVNTACSSSLVSLHLAVRALRNGECDM
                     ALAGGATVLSTPTVLVDFSRQRGLAPDGRCKAFADSADGTSWAEGAGMLLLQRLSDAR
                     REGRPVLAVIRGSAVNQDGASNGLTAPNGRAQRQVIEDALRDAGVGPDQVDAVEAHGT
                     GTELGDPIEAGALLATYGTARTAERPLWLGSLKSNIGHTQAAAGVAGVIKMVLAMRHG
                     RLPRTLHVDRPTTRVDWEKGGVRLLTEPVPWPGEAGEPRRAGVSSFGASGTNAHVVLE
                     SVPAGEPPAAGRPEDTGGAWTVSGRGPAALRAQAARLYDALTGTGTGTGQGAGQGAGP
                     GTAEVAGALAHARTAFRHRAVVLGGNRAELLAGLRELAEEEHPGPRVVTGTAPATERR
                     TAFLFSGQGSQRAGSGRGLYRRHPVFARALDEVCAALEPHLHRPLRDLMFAEPGSPEA
                     EPLDRTEFTQPALFALQTALFRLAEHHGLRAEALCGHSVGEIAAAHAAGVLTLPDAAR
                     LVAARGRLMQALPAGGAMAALRATAEEIAPLLERRAGELALAAVNGPSSVVVSGDEAA
                     VLELLEQWRAEGREARRLAVSHAFHSPRMDGMLTQFDRVARTLTFAPPTIPLVSTLTG
                     TPVTEETLCTADHWVRQAREPVRFLDAMRTLRADGIDTFVELGPDGVLSAMARDCADD
                     RPDGDTTGAGDGETPDPLLTLPLLRRSVPETGDAEHPGGFERALATAYAHGVPLRLAP
                     APDAASLAVAAELPTYAFQRTHYWLDAPAAPAALPAGLDDAGHPLLSAALDLPGGRGT
                     VWTGALSAATLPWAADHSVHGRTVLPGTALLDLALHAAPRVGELTFEAPLVLPEDGEV
                     RLRVVLAEPDASGVRELSVHSAGEDGGWTRHATAVLDTGTTTAGEPAGAPPAAWPPGD
                     AEPLDLAAEYERFADAGIGYGPAFRGLRSAWRDGDAILADVRLPGELAGEADRYGIHP
                     ALLDAALHTAAAALGGAHGMLPFTWNGVTLHARGAHAIRVRLTPAGPDAVAVTAVDPA
                     GRPVFTAASLTLRPVTTGQLTAAEAARAPLYRVRWTGLPDTGTARDHTWAVAGGPGDL
                     LPGETPHHPDLASALADTGTAPFRVLADLRGYGTATPRELASQALALVQQWADAAEAA
                     EGRLVLVTRRAVDIGDGVTDPAAATVWGLVRAAQSEHPGCFALLDTDDSPRSRQLLPR
                     VAGTAEQLALRDGTLLAPSLTRATLPAGARLPALDGTVLITGGTGSLGAEAARHLVTR
                     HGARRLLLTSRSGPQAPGAAELVAELAALGAHADVAACDVADRAALRALLDRVPAGHP
                     LTAVLHTAGVLDDGVLTAQTPQRLAAVLRPKADAVRNLHELTQGHALSAFILYSSAAG
                     VLGSAGQSGYAAANAYLDSFAVWRRSRGLPAVSLGWGPWDGGGMASGLGGTDTARLRR
                     SGIAPLSRAEGLAALDAALAAGGDDTAPAHLLPIRVDAVTLRGADTVPAVLRDLAGTA

  Query Match             52.1%;  Score 128.2;  DB 1;  Length 43280;
  Best Local Similarity   70.2%;  Pred. No. 7.6e-08;
  Matches  172;  Conservative    0;  Mismatches   73;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||| |||| |  ||     |  ||  |||| ||  | |||| |||  | ||||| |||||
Db      30648 CTGACCGGGCTCACGGCACAGGAAGGGCACGCCCTCGTCCTGGCCGAGATCCGCGCCCAC 30707

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
                ||  | || |||||||||  | ||  |||| | ||||||   ||||| || |||
Db      30708 GCGGCGGCGGTGCTGGGCCACGGCTCCGACGACTCGATCCCCGAGGACCGGGCCTTCAAG 30767

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||||||||||||||| |||||||||| ||  | ||||||||||   |    |||||
Db      30768 GACCTCGGCTTCGACTCGCTCACCGCCGTGGAGATGCGCAACCGGCTGAGCGCGGCCACC 30827

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || ||||| || ||| |||||||||| |||||||||||| | |||   |   ||  ||  |
Db      30828 GGCCTCCGGCTGCCCGCCACCCTCGTCTTCGACCACCCGACCCCGGGCGAGCTGGCCGGC 30887

Qy        241 CACCT 245
              |||||
Db      30888 CACCT 30892


RESULT 15
AE007061/c
LOCUS       AE007061               15970 bp    DNA     linear   BCT 27-APR-2001
DEFINITION  Mycobacterium tuberculosis CDC1551,  section 147 of 280 of the
            complete genome.
ACCESSION   AE007061 AE000516
VERSION     AE007061.1  GI:13881778
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis CDC1551.
  ORGANISM  Mycobacterium tuberculosis CDC1551
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 15970)
  AUTHORS   Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
            Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
            Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
            Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
            Gill,J., Mikula,A. and Bishai,W.
  TITLE     Whole genome comparison of Mycobacterium tuberculosis clinical and
            laboratory strains
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 15970)
  AUTHORS   Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
            Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
            Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
            Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
            Gill,J., Mikula,A. and Bishai,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .15970
                     /organism="Mycobacterium tuberculosis CDC1551"
                     /strain="CDC1551"
                     /db_xref="taxon:83331"
                     /note="clinical strain"
     gene            70. .726
                     /gene="MT2106"
     CDS             70. .726
                     /gene="MT2106"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK46385.1"
                     /db_xref="GI:13881779"
                     /translation="MRIAALVAVSLLIAGCSREVGGDVGQSQTIAPPAPAPSAAPSTP
                     PAAGAPITTIVSWIEAGHPVDPAAYHVATRDGVTTQLGDDVAFSASSGTVACMTDARH
                     TSGTLACLVRLANPPPRPETAYGEWKGGWVDFDGIHLQVGSARADPGPFVYGNGPELA
                     NGDTLSIGDYRCRSYQAGLFCVNYAHQSAVRFASAGIEPFGCLKPAPPPDGVGVAFGC
                     "
     gene            complement(763. .3327)
                     /gene="MT2107"
     CDS             complement(763. .3327)
                     /gene="MT2107"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAK46386.1"
                     /db_xref="GI:13881780"
                     /translation="MRIAVTGASGVLGRGLTARLLSQGHEVVGIARHRPDSWPSSADF
                     IAADIRDATAVESAMTGADVVAHCAWVRGRNDHINIDGTANVLKAMAETGTGRIVFTS
                     SGHQPRVEQMLADCGLEWVAVRCALIFGRNVDNWVQRLFALPVLPAGYADRVVQVVHS
                     DDAQRLLVRALLDTVIDSGPVNLAAPGELTFRRIAAALGRPMVPIGSPVLRRVTSFAE
                     LELLHSAPLMDVTLLRDRWGFQPAWNAEECLEDFTLAVRGRIGLGKRTFSLPWRLANI
                     QDLPAVDSPADDGVAPRLAGPEGANGEFDTPIDPRFPTYLATNLSEALPGPFSPSSAS
                     VTVRGLRAGGVGIAERLRPSGVIQREIAMRTVAVFAHRLYGAITSAHFMAATVPFAKP
                     ATIVSNSGFFGPSMASLPIFGAQRPPSESSRARRWLRTLRNIGVFGVNLVGLSAGSPR
                     DTDAYVADVDRLERLAFDNLATHDDRRLLSLILLARDHVVHGWVLASGSFMLCAAFNV
                     LLRGLCGRDTAPAAGPELVSARSVEAVQRLVAAARRDPVVIRLLAEPGERLDKLAVEA
                     PEFHSAVLAELTLIGHRGPAEVEMAATSYADNPELLVRMVAKTLRAVPAPQPPTPVIP
                     LRAKPVALLAARQLRDREVRRDRMVRAIWVLRALLREYGRRLTEAGVFDTPDDVFYLL
                     VDEIDALPADVSGLVARRRAEQRRLAGIVPPTVFSGSWEPSPSSAAALAAGDTLRGVG
                     VCGGRVRGRVRIVRPETIDDLQPGEILVAEVTDVGYTAAFCYAAAVVTELGGPMSHAA
                     VVAREFGFPCVVDAQGATRFLPPGALVEVDGATGEIHVVELASEDGPALPGSDLSR"
     gene            complement(3332. .15787)
                     /gene="MT2108"
     CDS             complement(3332. .15787)
                     /gene="MT2108"
                     /note="similar to PID:987099; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
```

FEATURES Location/Qualifiers

source 1. .43280
/organism="Streptomyces fradiae"
/db_xref="taxon:1906"

gene 1. .43280
/gene="tylG"

CDS 816. .14234
/gene="tylG"
/note="ORF1; putative"
/codon_start=1
/transl_table=11
/product="tylactone synthase starter module and modules 1 & 2"
/protein_id="AAB66504.1"
/db_xref="GI:2317860"
/translation="MSSALRRAVQSNCGYGDLMTSNTAAQNTGDQEDVDGPDSTHGGE
IAVVGMSCRLPGAAGVEEFWELLRSGRGMPTRQDDGTWRAALEDHAGFDAGFFGMNAR
QAAATDPQHRLMLELGWEALEDAGIVPGDLTGTDTGVFAGVASDDYAVLTRRSAVSAG
GYTATGLHRALAANRLSHFLGLRGPSLVVDSAQSASLVAVQLACESLRRGETSLAVAG
GVNLILTEESTTVMERMGALSPDGRCHTFDARANGYVRGEGGGAVVLKPLDAALADGD
RVYCVIKGGAVNNDGGGASLTTPDREAQEAVLRQAYRRAGVSTGAVRYVELHGTGTRA
GDPVEAAALGAVLGAGADSGRSTPLAVGSVKTNVGHLEGAAGIVGLIKATLCVRKGEL
VPSLNFSTPNPDIPLDDLRLRVQTERQEWNEEDDRPRVAGVSSFGMGGTNVHLVIAEA
PAAAGSSGAGGSGAGSGAGISAVSGVVPVVVSGRSRVVVREAAGRLAEVVEAGGVGLA
DVAVTMADRSRFGYRAVVLARGEAELAGRLRALAGGDPDAGVVTGAVLDGGVVVGAAP
GGAGAAGGAGAAGGAGGGGVVLVFPGQGTQWVGMGAGLLGSSEVFAASMRECARALSV
HVGWDLLEVVSGGAGLERVDVVQPVTWAVMVSLARYWQAMGVDVAAVVGHSQGEIAAA
TVAGALSLEDAAAVVALRAGLIGRYLAGRGAMAAVPLPAGEVEAGLAKWPGVEVAAVN
GPASTVVSGDRRAVAGYVAVCQAEGVQARLIPVDYASHSRHVEDLKGELERVLSGIRP
RSPRVPVCSTVAGEQPGEPVFDAGYWFRNLRNRVEFSAVVGGLLEEGHRRFIEVSAHP
VLVHAIEQTAEAADRSVHATGTLRRQDDSPHRLLTSTAEAWAHGATLTWDPALPPGHL
TTLPTYPFNHHHYWLDTIDGGGGDDATQEKESGPLTRELRGLPSSQKQLGFLLDLVCR
HTAVVLGLDTAAEVDPDLSFKKQGIQSMTGVELRNRLLTETGLALPTTLVYDRPTPRA
LAQFLHTELLDGSPSGSVLAPAQKSFEAQEPIAVVGMGCRFPGGVGSPEALWRLVVEG
VDAVSPFPGDRGWDVEGLYDPEPGVAGKSYVREGGFLHDAAEFDAEFFGISPREAVAM
DPQQRLLLETSWEAIERAGIDPHSLHGSRTGVYAGVMPQEYGPRLAEGAEGSDGYLLT
GTSGSVVSGRVAYTLGLEGPAVTVDTACSSSLVALHLAVQALRGGECDMALAGGVTVM
AGPGMFVEFSRQRGLAADGRCKAFADGADGTAWAEGAGVVLVERLSDARRLGHPVLAV
VCGSAVNQDGASNGLTAPSGPSQERVIRQALGNARLTVADVDVVEAHGTGTRLGDPIE
AQALLGTYGRDRDGGRPVWLGSLKSNIGHAQAAAGVAGVIKMVLAMRYGWLPRTLHVD
EPSRHVDWSAGGVWLLTEAREWPGVDRPRRAAVSAFGVSGTNAHLILEAPDTAEAESA
TTPVRSEVSESAAVLDARSGVVPVVVSGRSRVVVREAAGRLAEVVEAGGVGLADVAVT
MAGRSRFGYRAVVLARGEAELAGRLRALAGGDPDAGVVTGAVVDPETGSGGGGVVLVF
PGQGTQWVGMGAGLLGSSEVFAASMRECARALSVHVEWDLLEVVSGGAGLERVDVVQP
VTWAVMVSLARYWQAMGVDVAAVVGHSQGEIAAATVAGALSLEDAAAVVALRAGLIGR
YLAGRGAMAAVPLPAGEVEAGLAKWPGVEVAAVNGPASTVVSGDRRAVAGYVAVCQAE
GVQARLIPVDYASHSRHVEDLKGELERVLSGIRPRSPRVPVCSTVAGEQPGEPVFDAG
YWFRNLRNRVEFSAVVGGLLEEGHRRFIEVSAHPVLVHAIEQTAEAADRSVHATGTLR
RQDDSPHRLLTSTAEAWAHGATLTWDPALPPGHLTTLPTYPFNHHHYWLDTTPTTPAT
TTQSPTDAWRYRVTWKALTESSPVRPHSIGRCLLVAPPTTDGELLDGLTTVLSERGAS
VARLEVPIGARRAEVAELLKPSMESAGEENTTVVSLLGLVPSTDAVRTSIALLQAVSD
IGVPAARVWALTRRAVAVVPGETPQDAGAQLWGFGRVAALELPDIWGGLIDLPETAEL
TRTPETSQPPQTPERLPQTPNRRALELAAAVLAGRDGEDQVAVRASGIYGRRVSRAAA
AGAASWQPSGTVLITGGMGAIGRRLARRLAAEGAERLVLTSRRGPEAPGAAELAEELR
GHGCEVVHAACDVAERDALAALVTAYPPNAVFHTAGILDDAVIDTLSPESFETVRGAK
VCGAELLHQLTADIKGLDAFVLFSSVTGTWGNAGQGAYAAANAALDALAERRRAAGLP
ATSVAWGLWGGGGMAAGAGEESLSRRGLRAMDPDAAVDALLGAMGRNDVCVTVVDVDW
ERFAPATNAIRPGRLFDTVPEAREALTAAGTTSATPDGAPELARRLSMLNETERLRKL
VELVRTEAAFVLRHPNTDAIGAERPFKSAGFDSLTSLELRNRLNAGTGLKLPATVIFD
HPSPTALARLLLDRLTGAGAPAPAADEPPLPVAVADDDPVVIVGMACRFPGGAGTPEA
LWKLVTEERDVIGAAPTDRGWDLDSVYDPEPGVAGKTYVREGGFLHDAAEFDAEFFGI
SPREAVAMDPQQRLLLETSWEAIERAGIDPHSLHGSRTGVYVGLTHQEYASRLHEAPE
EYEGYLLTGKSASVVSGRISYTLGLEGPSLSIDTACSSSLVALHNAAQALRGGECDMA
LAGGVTVMAAPGLFVEFSRQRGLAADGRCKAFADGADGTAWAEGAGVVLVERLSDARR
LGHPVLAVVCGSAVNQDGASNGLTAPSGPSQERVIRQALANARLTVADVDVVEAHGTG
TRLGDPIEAQALLGTYGRDRDAECPVWLGSLKSNIGHAQAAAGVAGVIKMVLAMRYGW
LPRTLHVDEPSRHVDWSAGGVRLLTEAREWPGVDRPRRAAVSAFGVSGTNAHLILEAP
EALEALEATDAPEAPEAPEAPDVTDVTEALEAPDATEAEGAKAPGSPEEAQPAVGVVP
VVVSGRSRVVVREAAGRLAEVVEAGGVGLADVAVTMAGRSRFGYRAVVLARGEAELAG
RLRALAGGDPDAGVVTGAVVDPETGSGGGGVVLVFPGQGTQWVGMGAGLLGSSEVFAA
SMRECARALSVHVEWDLLEVVSGGAGLERVDVVQPVTWAVMVSLARYWQAMGVDVAAV
VGHSQGEIAAATVAGALSLEDAAAVVALRAGLIGRYLAGRGAMAAVPLPAGEVEAGLA
KWPGVQVAAVNGPASTVVSGDRRAVAGYVAVCQAEGVQARLIPVDYASHSRHVEDLKG
ELERVLSGIRPRSPRVPVCSTVAGEQPGEPVFDAGYWFRNLRNRVEFSAVVGGLLEQG
HRRFIEVSAHPVLVHAIEQTAEAADRSVHATGTLRRQDDSPHRLLTSTAEAWAHGATL
TWDPALPPGHLTTLPTYPFNHHHYWAVTSPAGVGDAAAGRFGMTWEDHPFLRGGLPLA
DSGERVFAGRLAGSEHDWLTDHAVSGVTLLPGTAFVEFALHAGAATGCGRLEELSVEA
PLVLPAAGGVRVQMRVSAADESGRRRVAIHSAPEAAVHSAAEGGDSAGVWTRHGEGTL
VPDPEPTPPDADWARAWPPAGERVEPAELYERFGALGYEYGEAFAGVRAVWRQPDALL
AEVLLPDRASTGAGRFGVHPALLDAALQPWIAGGLLEVPEDAVLLPFAWQGVSLYATG
AGALRVRLTKAGDGAVSLQAADTSGAAVLSLGALVMRPLARRKLDVLLGTDAGERSLY
RVEWQPRLLPAGPPRSWAVLGPDADRLAGTPGLGDQPDGGPTALYPEVRALRKALAAG
APRPEAVVLPVLSGAGATPESVRQTTERCLTALQDWLDAEELVDTPLIVLTRGAVAAV
PGEEIGDLACAGVWGLVRSARSEHPGRFALVDTDGHPDDRTALPLALRAVLDGAGQLS
LRAGTARTPVLLRAGTPEEQRGPAFDPAGTVLVTGATGTLGRLLARHLAAEHGVRHLL
LLSRGGRAAEGADELAAELAGLEAEPCFAACDAADREALARVLAEVPADRPLTGVIHA
AGVLDDGTLDALTPERIGTVMRPKADAALNLHELTRTSPLSVFAVFSGAAGILGRPGQ
ANYAAANTFLDALAQHRRAHGLPAVSLAWGLWGGATGMTGHLSGTDLRRMRRSGIAPM
THDQGLALFDRALAASAEDPLLVPMRLDLAALVRERAEHGPDAVPGPLLGLLPARAAV
RQAAAPVRGGAPAPAGGEGTAERLAGLGEEARLRELVRLVRAEVSGVLGYSGPDAVEP
GRPFKDLGFDSLTAVELRNRLGAATGLRLPTALVFDRPTSQAVAEYLAAELAGPRDGG
DTAAAAFEGLEALAAAVGALAEDDLRRDVLRRRLTELAAALTPQGRNPSAPAPAPSDL
DERLDSANDDDLFAFIEEQL"

misc_feature 942. .3929
/gene="tylG"
/function="encodes starter module"

misc_feature 942. .2156
/gene="tylG"
/function="encodes beta-ketoacyl ACP synthase KS-Q of starter module"

misc_feature 2571. .3557
/gene="tylG"
/function="encodes acyltransferase domain of starter module"

misc_feature 3675. .3929
/gene="tylG"
/function="encodes acyl carrier protein domain of starter module"

misc_feature 3993. .8471
/gene="tylG"
/function="encodes module 1"

misc_feature 3993. .5264
/gene="tylG"
/function="encodes beta-ketoacyl ACP synthase of module 1"

misc_feature 5631. .6617
/gene="tylG"
/function="encodes acyltransferase domain of module 1"

misc_feature 7410. .7949
/gene="tylG"
/function="encodes beta-ketoreductase domain of module 1"

misc_feature 8220. .8471
/gene="tylG"
/function="encodes acyl carrier protein domain of module 1"

misc_feature 8541. .13970
/gene="tylG"
/function="encodes module 2"

misc_feature 8541. .9812
/gene="tylG"
/function="encodes beta-ketoacyl ACP synthase of module 2"

misc_feature 10260. .11246
/gene="tylG"
/function="encodes acyltransferase domain of module 2"

misc_feature 11319. .11876
/gene="tylG"
/function="encodes dehydratase domain of module 2"

misc_feature 12861. .13415
/gene="tylG"
/function="encodes beta-ketoreductase domain of module 2"

misc_feature 13719. .13970
/gene="tylG"
/function="encodes acyl carrier protein domain of module 2"

CDS 14351. .19945
/gene="tylG"
/note="ORF2"

```
                     DRSWLRHATGVLSSAEAPSAAKLGAWPPAGAEAVDLTAFYEGMAAAGLDYGPVFQGLR
                     SVWRDGDDVFAEVELPDESAAEAAAFALHPALLDAALHALAAGGLVSLDDSPALPFAW
                     SGVVVQAAGAAMVRVKLSRTATGSVTLAVADGAGDPVASVESLSLRAASAEQLPGAVG
                     GDQLFGLEWTRFTLPSATDALMIETVADFDALRESEGTPDVVVVPVVPCPAGSGRETA
                     DRVHSTTSEVLALVQWWLAEERAGRLALVTRPGDLAHAAVWGLVRSAQSENPDRIVLV
                     EAEETDEAVRVLPAALASGEPQFAVRDGEVFVPRLVKATRSATGTATGTPDFGVGPVL
                     LTGASGALGGLVARHLVAEHGVRSLLLLSRRGAEAPGAVELEAELAAWGAEVRWAACD
                     VADREAVSGMLNGLGEQSLSAVVHTAGVLGDGIVASLTPERMREVFRPKVDAVLNLHE
                     CTRDMGLAAFVVFSSVAGMVGSAGQASYAAANSFLDAFSAHRRREGLPAISLAWGVWE
                     QSGAMTDGLVEADRARMARSGVLPLPPEEGLRLFDAALASDEAVLAPVRIDTGALRAG
                     EAPPVLRALVPAAAARRTARAAAAPSASSLAERLEEMPEAERERKTVLDLVRAEVAAGL
                     GHASDRTVRPEHAFQDLGFDSLTAVELRNRLNKATGLRLPATLVFDHPTPALLARHVF
                     MAAAGAAEPALVDSLLADLDRVEQELVTTLAASEARDRILSKLQGLLAIAGEFGKAPA
                     QEGPGTGGSDLESATDDEVFDLLGKEFGIS"
     gene            21273. .31847
                     /gene="pteA2"
     CDS             21273. .31847
                     /gene="pteA2"
                     /codon_start=1
                     /transl_table=11
                     /product="modular polyketide synthase"
                     /protein_id="BAB69304.1"
                     /db_xref="GI:15824140"
                     /translation="MSCRYPGGVDSPEDLWELVAGGRDAVSEFPADRGWDLDALYDPD
                     PESRGTSYAREGAFLRDLARFDAGLFGISPREALAMDPQQRLLLETSWELFERAGIDA
                     ASLAGSRTGVFAGVMNHEFLAALQNSPQDLEGYLGTGTSGSVASGRVAYTFGLEGPAV
                     TVDTACSSSLVALHLAVQSLRNGECSLAVAGGVTAMVGPATFVGFSRQRGLAPDGRCK
                     PFAAGADGTGWGEGAGLLLVERLSDARRNGHEVLAVVRGSAVNQDGASNGLTAPNGPS
                     QQRVIRQALAAAQLTTAQVDAVEAHGTGTRLGDPIEAQALLATYGQGRPDDKPLWLGS
                     IKSNIGHTQAAAGVAGIIKMVMAMRHGELPRTLHVDAPSVHVDWTAGAVRLLTERTPW
                     SASDQPRRAGISSFGISGTNAHTIIEEAPSAEAFAQDTGAPEGPEEGAESAEPVASEA
                     ARPAAGGPVPWVISGRTEEALRAQAERLLSRTDAEVPPLDVAFSLATTRTALEHRAVL
                     QGETPAELAEGLRAVASGSPARGVARGLSGTGGRVGFLFSGQGSQRLGMGRELYASYP
                     AFAAAYDEVCAQLDAVVDVDSEELHRTGAAQPALFAVEVALFRLLESWGIRADYVAGH
                     SVGELAAAHVAGVLSLDDAAKLVSARARLMQALPAGGAMVAVQATEDEVLPHLTDAVG
                     IAAINGPRSVVVSGAEDAVTAIAEAFRQQGRKSSRLKVSHAFHSPLMDPILEEFAEVV
                     RSLTFHRPQIPVVSNLTGRLAEPYTPQYWVRHVREAVRFADGVRTLNDLGVTTFVEIG
                     PGGVLSALAEGCLDDDVVTVPALRADRPERVALVAAVSELHAHGVSPDWHAFFPGARR
                     VDLPTYAFQYERYWLDASPAAPGGVRAAGLGSADHPLLGAAVSLAGGDERLLTGRLSL
                     RTHPWLSDHAVLGTALLPGTAFVELAVRAADEAGCALLEDLTLEAPLVVPEQGGVAVQ
                     VWVGAVDGTGRRPLTVHSRPEDDSDLPWVRHATGFVTEGSEEAGAAAADRTLTAWPPT

  Query Match             53.2%;  Score 130.8;  DB 1;  Length 78210;
  Best Local Similarity   70.7%;  Pred. No. 3.1e-08;
  Matches  174;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || |||||  | | |   |||  | |   ||    |||||||  |  ||||||||   | |
Db      26139 CTCGCCGGTCTGGCCGCCGACGAGCGCGACGAGGCCCTCCTGGAACTGGTCCGGGGCGAG 26198

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 120
               |||||  |||| || || ||| || |  ||     || |||  |||      ||| ||||
Db      26199 GTCGCCGCCGTGCTCGGTCACTCCTCGTCCCGGGCCGTCCAGCCCACGTACGCCTTCCAG 26258

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||||||||||||||| || || ||||| || || ||||||||||| ||    |||||
Db      26259 GACCTCGGCTTCGACTCGCTGACGGCCGTGGAGCTGCGCAACCGGCTCGCCACGGCCACC 26318

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || || ||||| ||| ||||||||| |||||||||||||| |||| || ||||| |||
Db      26319 GGCCTGCGCCTGCCCGCCACCCTCGTCTTCGACCACCCCACCCCCGCCGCCCTCGCCCGG 26378

Qy        241 CACCTC 246
              |||  ||
Db      26379 CACATC 26384


RESULT 13
AX211705
LOCUS       AX211705               65140 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 1 from Patent WO0159126.
ACCESSION   AX211705
VERSION     AX211705.1  GI:15523937
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 65140)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 1 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .65140
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT     8270 a  25171 c  22273 g   9426 t
ORIGIN

  Query Match             52.8%;  Score 129.8;  DB 6;  Length 65140;
  Best Local Similarity   70.6%;  Pred. No. 4.3e-08;
  Matches  173;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || | | | |      ||||  ||||  |||     | || |||| | ||||||||  | |
Db      26732 CTCGGCCGACTCGACCACGCCCAACGCCACGAGGCACTGCTCGACATGGTCCGCAGCAGC 26791

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 120
               |||   |||||| |||||| ||  |  | || |||||  | |||| ||||| ||||||
Db      26792 GCCGCGCTCGTCCTCGGCCACGCCGACGGCAACGCCATCGACGCCGAACGCGCCTTCCGC 26851

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||||||||||||||| |||||||| |||||||| |||||||  ||   |  | |||||
Db      26852 GACCTCGGCTTCGACTCGCTCACCGCGGTCGAACTCCGCAACCGTCTGCGCACCGCCACC 26911

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || || | |||  | ||||| | | ||||||||||||||| ||   || ||||| | |
Db      26912 GGCCTGCACCTGTCGGCCACCATGGTCTTCGACCACCCCACCCTGTCCGCCCTCGCGGAG 26971

Qy        241 CACCT 245
              |||||
Db      26972 CACCT 26976


RESULT 14
SFU78289
LOCUS       SFU78289               43280 bp    DNA     linear   BCT 13-AUG-1997
DEFINITION  Streptomyces fradiae tylactone synthase, starter module and modules
            1-7, (tylG) gene, complete cds.
ACCESSION   U78289
VERSION     U78289.1  GI:2317859
KEYWORDS    .
SOURCE      Streptomyces fradiae.
  ORGANISM  Streptomyces fradiae
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 43280)
  AUTHORS   DeHoff,B.S., Sutton,K.L. and Rosteck,P.R. Jr.
  TITLE     Sequence of Streptomyces fradiae tylactone synthase gene tylG
  JOURNAL   Unpublished (1996)
REFERENCE   2  (bases 1 to 43280)
  AUTHORS   DeHoff,B.S., Sutton,K.L. and Rosteck,P.R. Jr.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-1996) Eli Lilly and Company, Lilly Corporate
            Center, Indianapolis, IN 46285, USA
COMMENT     tylG.
```

```
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
  TITLE     Genome sequence of an industrial microorganism Streptomyces
            avermitilis: deducing the ability of producing secondary
            metabolites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
  MEDLINE   21477403
REFERENCE   2  (bases 1 to 78210)
  AUTHORS   Ikeda,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .78210
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
                     /note="Actinobacteridae"
     CDS             complement(1. .1254)
                     /codon_start=1
                     /transl_table=11
                     /product="putative lipoprotein"
                     /protein_id="BAB69301.1"
                     /db_xref="GI:15824137"
                     /translation="MGEALITLCTRSTRLAATAGVLGGVLAVTALGGTSNAATSGNGA
                     GNHPTSQARINQAGAQEASDARIEITPGQGTYSVGTNHPVSVTVTSGKLTKVTMTAVA
                     TGAEIPGTLSTDGTSWKPNGRLERATKYQIAAEAGDVKGRSTTGSATITTVSPANDFI
                     GHLSPEDGSTVGVAMPVTVNFDKAISDKAAVQSEIRVSSSSGQRVVGHWFNDHRVDFR
                     PENYWRPGSTVTVTLNRHGIQKTVTFKIGRSQISTVDAKTKQMTVVRDGKTIKTIPIS
                     SGSPEHPTYNGQMVISEKFRQIHMNGASVGLTEKNGKPSYDIKAVPHAMRLTDSGTFI
                     HGNYWGTDSVFGEANTSHGCVGLKDVRGDSHGRQPAAWFFDHSIVGDVVIVKNSEEKT
                     MLAPDNGLSDWNMSWSEWVAGSPAT"
     CDS             complement(1256. .2290)
                     /codon_start=1
                     /transl_table=11
                     /product="putative transposase"
                     /protein_id="BAB69302.1"
                     /db_xref="GI:15824138"
                     /translation="MPVEFLSHDQVSRYGRFASAPTPGELEQFFRLDAKALESARSKR
                     APANRLGWAVQWGCVRMLGVFPTEDLSVVPEVVVRFAAEQLGVDPGEFAAYGARRQNR
                     YEHAWEIRDAYAYRESPAAEAQVREFLATRVWASWEGPRALFDRAVGWLVDNRVLLPG
                     ITTLTRLVAEVRAQENAALYRTLDEAVPEKLRQSMRDLLKVPEGRRVSELERLRRSPM
                     GVSGSADRLAFSERIRMEAVAMFTAGPGSTDIAKELRVSVQSVQRWRSGLAGDRPGRG
                     AVPQPVSVELSETLFAALDEERAKGFAPSWTIRYLHHDFPIFGQRYAYKWSQGCLHSF
                     GRSAQGYRDQ"
     gene            2935. .21081
                     /gene="pteA1"
     CDS             2935. .21081
                     /gene="pteA1"
                     /codon_start=1
                     /transl_table=11
                     /product="modular polyketide synthase"
                     /protein_id="BAB69303.1"
                     /db_xref="GI:15824139"
                     /translation="MVESELNATGNRDCSASADSADGIAVIGLACRLPGAADPAAFWR
                     LLSEGADAITDVPPDRWDGAVVADASAPDRTGIRRGGFLDSIGDFDAGFFGLSPKEAA
                     AMDPQQRLVLELAWEALEDAHVRAATLRSTRTGVFVGAMWDDYATLHHRSGLTAISPH
                     TVTGLHRSIIANRVSYFLGLNGPSLTVDSGQSSSLVSVHLACESLRKGESTIALAGGV
                     NLNIVPESALGAAKFGGLSPDGRCFTFDARANGYVRGEGGGIVVLKPLARAVADGDPV
                     YCVIRGSAVNNDGGGDGLTVPLQSGQEQVLRLAYERAGVNPADVGYVELHGTGTKVGD
                     PIEAAALGAVLGTGREPGRPLRVGSAKTNVGHLEGAAGITGLLKAVLSLNHRELPASL
                     NFGTPNPAIPLDRLNLRVQTEHSEWAAEDGRPLLAGVSSFGMGGTNCHIVLASGAAVE
                     AGEGATAGGGAAPTFWPLSAKTASGLRAQAAALLAHTEAPLGLALPDVGWSLATGRTA
                     FEHRAVVVGEVRDDFLRALRELSTGGIDAALTTGRAGPRGELAFLFSGQGSQRAGMGR
                     ELAESVPAFASALDEVCAHLDPLLPRPLREVMFAPVGTGEAAELDQTLYTQTSLFAVE
                     VALFRLLESWDITPDVLMGHSIGELAAAHVAGVLSLADACALVAARGRLMQALPAGGA
                     MIAVQAAEDEVRARIGERTDRVSIAALNGPDSVVVSGDEDLATEIADAFAALGRKTSR
                     LRVSHAFHSPHMEPMLAEFRNVAEGLTFHAPRIPIVSNVTGRLSEEPSGYEGCTAAYW
                     VRHVREAVRFADGVARLDEQGVRTYLELGPGGVLASMARAGADSESLFVPALRTRRPE
                     AQALLAAVASLHVHGVEPDWDALFGGGRGARRKVALPTYAFERQRYWLDGDTTSPVAT
                     TGWAPVAPQPDAIAETDSAPVGALGRRLAGLAESERDEAVLDMVRAHIAAVLGHAEAR
                     QVETEWTFKDLGFDSLSSVELRNQLSEATGLRLPSGLLFDHPTPAALARHLAAQSLGG
                     GDAAAAAPATVADPHEPIAIVAMSCRLPGGVSSPEDLWRLLASGGDGISGFPTDRGWD
                     VETLYDPEPGTPGKSSTRHGGFLHEAAEFDAGFFGISPREAAAIDPQQRLVLETAWEA
                     FERAGIDPATLRGSRAGVFIGATAQDYGPRLHEPADGVEGYLLTGTTTSVASGRVAYS
                     FGLEGPAVTVDTACSSSLVALHLAVQSLRSGECTMALAGGVTVMSTPGMFVEFSRQRG
                     LSADGRCKAFAAGADGTGWAEGVGMLLVERLSDARRNGHQVLAVVRGSAVNQDGASNG
                     LTAPNGPSQQRVVRQALAAAGLAPADVDAVEAHGTGTLLGDPIEAEALIAAYGQGRDG
                     DRPLWLGSLKSNIGHTQAAAGVAGVIKMVMAMRHGVLPRTLHVDEPTPHVDWSASGVR
                     LLTEAVEWPARDLPRRAAVSSFGISGTNAHVIVEQAATSEEPARGTAVAGPVPWVVSA
                     KSDVALREQAGRLLSFLDAEGVGGAGVRPEDVALSLATSRAVLERRAAVVGGDLAELR
                     QGLEALASGASGAAPVGGAVGGRVGFLFSGQGSQRIGMGRELYAAYPVFAAAYDEVCA
                     FLDAPVDVDSEELNQTGTTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVA
                     GVLSLEDAAKLVSARAALMQALPDGGAMVAVQATEDEVLRHLTDGVGIAAVNGPRSVV
                     VSGAEDAVTAVGEIFAGRGRKTSRLKVSHAFHSPLMDPMLEEFAAVVRGLTFSESRIP
                     VVSNLTGRLAEVYTPEYWVRHVREAVRFADGIETLGELGVTTFVEVGPGGVLSALAQG
                     CFDDGTDIVTVPVLRADRPEPYALVTAIGRLHTHGISPDWQVLFPGARRVDLPTYAFQ
                     RERYWLDAPRAATAGAPDAMDAEFWDSVESEDRASLGALLGLEPTELDVVASKLSAWR
                     RKRREQSTTDGWRYRITWQPLDDMVRAAPSGTWLYVVSEETAWTEAIRAGLTELGVAL
                     VPFAVGEDTDRDALARAFADAGQEQADGILFAAAQGEAGANALQRLVLLVQALGDAGI
                     EAPLWCLTSGAVSTGTADLLTDPVSAQLWGLGRIVALEQPQRWGGLVDLPAELDSRAL
                     ERLAGLLAQADEDQLAVRASGVYGRRLVRAPLSAAPADAPWSPRGTVLVTGGTGALGG
                     QVARWLAGAGAEHLVLTSRRGPDAPGAAELKAELEELGAQVTVVACDAADRDALAELF
                     GRYIVNAVVHTAGVLDDGLVESLTPERLNHVLRPKVDAALHLHELTRDRQDLDAFVLF
                     SSMTGVWGNGGQGAYGAANAFLDALAEHRRAQGVPALAVAWGSWADGGMADGAAGDHL
                     RRRGVRAIAALPAISVLHGALTHGETSMTVADVDWERFVPAFVGTRPCPLLLGVPEAR
                     EALEARARAAQSPEVPVSALVQRLLGATPGEQGRILLDLVREQAATVLGHTGKGAFEA
                     DRTFRETGFDSLTAVELRHRLNTATGLRLPTTLVFDHPTPTALARHLGEELLGRHGLE
                     TAAAQATAAAAADEPLAIVGMSCRLPGGVQSPEDLWELLSSGRDAMSGFPTDRGWDID
                     ALYDPDPDRPGKTYSREGGFIQGADRFDAGLFGISPREALAMDPQQRLLLETAWEAFE
                     RAGIALASVRASRTGVFIGTNGQDYANGLRNAPEEIEGYALTGKAASVVSGRISYTFG
                     LEGPAVTVDTACSSSLVALHLAAQALRSGECSMALVGGVTVMTTPDLFVEFSRQRGLS
                     PDGRCKAFAAGADGTGWGEGVGLLLVERLSDARRNGHQVLALVRGSAVNQDGASNGLT
                     APNGPSQQRVIRQALASAGLTPADVDAVEAHGTGTRLGDPIEAQALLATYGQEHSDDR
                     PVWLGSVKSNIGHTQAAAGVAGVIKMVLAMRHGVLPQTLHVDEPTAHVDWSAGAVRLL
                     TEPVQWPGTTGPRRAAVSSFGIGGTNAHTIIEEAPAAAAAAEPAREHRPVPVPWVLSAK
                     SEVALRAQAERLLAFATDDVSPVDAGFSSATTRSALEHRAAVIGSDPAELRAGLEALA
                     AGEPAANVVAGRAHSADKVGPLFSGQGSQRLGMGRELYAAYPVFAAAYDEVCALLDAT
                     VDVDAETLHQTGSTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVAGVLSL
                     EDAAKLVSARAALMKALPAGGVMVAVQAAEDEMLPHLIDRAGIAAINGPRSVVISGAE
                     DAVTAIAEVFAAQGRKTSRLKVSHAFHSPLMDPMLEEFAEVVRGLTFSEPWIPVVSNL
                     TGRLTESYTPEYWVRHVREAVRFADGVLTLHELGVTTFVEIGPGGVLSALVQGCLDDG
                     TDIVTIPALRADRPEPQALTTAYAQLHVSGAEVDWHAFFPGARRVELPTYAFQRERYW
                     LDAPASAGDMRAVGQGEAGHPLLGAAVPLADSVGHLLTGRLSKHTHPWLVDHAVSGVA
                     LLPGTAFVELAITAADAVGCDLLEELTLETPLLVPERGGVALQVHVGADDGSGRRPLT
                     VHSQIDDGDAYAHEDDSERAWVRHASGVLTVATGGPVDGSLTAWPPVGAEPVDVDGFY
                     DRLAGMALDYGPVFQGLRAAWRSGDDFFAEVELPGAERTDAGAYGLHPALFDAALHTV
                     WLGAVEPEAGTGNGLLPFAWSGVRLAAAGASDLRVKVSRAGTGTSTVSLVLADGTGGP
                     VAEVDSLTLRPVPADQLRAGSGSDDAVFGLKWTPVSLPSAPDGMRIETYEDLAALRAT
                     DPPADTEADAAAAADAVVVPCPTGTGADLATRVREVTHAVLELVQWWLAEERPGRLVL
                     VTRGGDLAQSAVCGLVRSAQTENPDRIVLVETSSDAHAGAGSGVDDADLARVLPGLIA
                     SGEPQAAVHADDEVRVPRLVRVAVPDAEPELPGLGSGTVLLTGASGGLGGLFARHLVA
                     EHGVRSLLLVSRRGGDAPGAAELTADLTAQGADVTWAACDVADRAAVRALLAGPGQSL
                     SAIIHTAGVLDDGIIGSLTPERLDAVFRPKVDAALNLHDLAGELSGDLSAFVLFSSVA
                     GTLGTPGQANYAAANTFLDALAEHRRAAGQRATSLAWGLWAQDSESAMTGSLDHTDLT
                     RIQRMGLAAIPSSDGLRLFDAALALDRAVVAPLRLDLGGLQGTVPPVLRGLRGTVRGT
                     ARRTAQTAPKGSLGQRLAGLAEAERERQVVLDLVRTEVAAVLGHAGAQSVGADDSFKDI
                     GFDSLTAVELRNRLNSAVGMRLPATLIFDYPNPLTLAHFLTGEAAGGEAATAVPTTLS
                     TATGTADEPIAIVGMACRYPGGVQSPEDLWQLVFSGRDAVSGFPEGRGWDVEQLYDPN
                     PDQWGTSYTREGGFLHDAADFDAEFFGISPREALAMDPQQRLLLETSWEAFERAGIDP
                     ATVRGSRTGVFAGVMYHDYGGRVHTSPPGLEGYLVNGSAGSIASGRVSYTFGLEGPAV
                     TVDTACSSSLVALHLAAQALRSGECSMALVGGVTVMAGPSVFVEFSRQRGLSADGRCK
                     AFGAGADGTGWAEGVGMLLVERLSDARRNGHKVLAVVRGTAVNQDGASNGLTAPNGPS
                     QQRVIRQALANAGLSSDEVDAVEAHGTGTRLGDPIEAQALIATYGQERSEDRPLWLGS
                     LKSNIGHSQAAAGVGGVIKMVMAMQHGVLPQTLHVDEPTPHVDWSAGAVQLLTEAVEW
                     PASDGPRRAAVSSFGVSGTNAHVIVEQAATSEVTARGTAVAGPVPWVLSAKSDAALRE
                     QAGRLLSFLDAEGVGGAGVRPEDVAYSLATSRAVLECRAAVVGEELGEFRQGLEGIVS
                     GAAPVGGAVGGRVGFLFSGQGSQRIGMGRELYAAYPVFAAAYDEVCALLDATVDVDAE
                     TLHQTGSTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVAGVLSLEDAAKL
                     VSARAALMQALPAGGAMVAIQATEAEVLPHLTHGVGLAAINGPRSVVLSGSEDAVLDI
                     AEVFAQQGRKTSRLKVSHAFHSPLMDPMLEEFAEVVRGLTFSEPRIPVVSNLTGRLAE
                     VYIPEYWVRHVREAVRFADGLHTLHELGVTMFVEIGPGGVLSALVQGCLDDGTHIVTV
                     PALRADRPEPYALVTAVGQLHTHGISPEWQALFPGARRVDLPTYAFQRERYWLDAPQA
                     SGDVRAAGLGAADHPLLGAAIATADSDGAVLTGLLSLGTHPWLADHAVNGIVLLPGTA
                     FVELAIRAGDETGCGWIEDLTLELPLVVPERGGVTLQVAVGPEDESGRRQLSVHSRVD
```

```
                QVALVLGHATGGDVDAGRAFRDLGFDSLTAVELRNRLNTVTGLRLPATLVFDYPTVRH
                LATYVLDELLGTDAEVATVQPAAVAVADDPIVIVGMACRYPGGVSSPEDLWRVLTEGT
                DAVSGFPTNRGWDVESLYHPDPDHPGTSYTRSGGFLHEAGEFDPGFFGMSPREALATD
                SQQRLLLESSWEAIERAGIDPVSLRGSRTGVFAGVMYSDYSAMLASPEFEGFQGSGSS
                PSLASGRVAYTLGLEGPAVTVDTACSSSLVAMHWAMQALRSGECGLALAGGVTVMSTP
                AVFVDFARQRGLSPDGRCKAFADAADGVGWSEGVGVLVLERQSDAVRNGHEILAVVRG
                SAVNQDGASNGLTAPNGPSQQRVIRQALASGGLTAGDVDVVEAHGTGTTLGDPIEAQA
                LLATYGRDREPERPLLLGSVKSNLGHTQAAAGVAGVIKMVLAMRHGVVPRTLHVDAPS
                SHVDWSEGAVELLSEQAAWPETGRVRRAGVSSFGISGTNVHVIVEQAPGAKAIAAAGA
                ARRTPGAVPVLLSGRGRSALRGQAARLLGHLQARPDAELVDVALSLATTRSRFEQRAA
                VVAQDRDQLIASLGALAADRPDPAVVEGEAAGRGRTAVLFTGQGSQRAAMGRELHEVQ
                PEFAAAFDAVCAVFDPLLDRPLREVVFAEDGSDEAALLDETGWTQPALFAVEVALFRL
                VESWGVRPDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARATLMQALPTGGAMIAIQA
                TEDEIAAHLDDTVAIAAVNGPQSVVISGDEEAAETIAATFAERGRKTKRLRVSHAFHS
                PRMDGMLDAFRIVAEGLTYRAPRIPLVSDLTGRRADDAEVCTAEYWVRHVREAVRFAD
                CVRTLRDAGATTFLELGSDGLLTAMAEDTLGDDHDAELVPMLRAGRAEELAAATALAR
                LQVRGVDVDWAAYLAGTGARRTDLPTYAFQHAYYWPQLPTPAAALAAADPADQQLWAA
                VERGDARELADILGLGEQDLTPLDSLLPALTSWRRGNQEKHLLDTLRYRVEWTRLSKP
                TAPVLDGTWLLVASDATAADQPALLDGLADALGSHGARVRRLLLDDSCADRAVLAERL
                ARTADVDAATQVLSVLPLDERDADDCPPLTRGLALTVALVQALADTGAQGRLWTATRG
                AVSTNPADPVTHPVQAAAWGLGRGVALEHPRLWGGLVDLPQVFDERAGQRLAGILAVK
                DAPDGEDQVALRATGVSGRRLVRHTVEALPTAAEFTATGTVLITGGTGGLGAEVARWL
                ARAGAQHLVLTSRRGPDAPGAAELRAELEGYGPSVSVVACDVADRDALAAVLTALPEE
                LPLTGVVHTAGVGHYGPLDTLSTAEFAGLTAAKLAGAAHLDALLADRELDFFVLFGSI
                AGVWGSGNQSAYGAANAYLDALALHRRARGLAATSVAWGPWAEAGMAADDAVSETLRR
                QGLGLLDPAPAMTELRRAVVRQDVTVTVADVDWQRYAPLFTSARPSALIAGLPEVRAL
                AADERTEQDATGASEVVTRVRALAEPEQLRLLTDLVRTESATVLGHSSADAVPEGRAF
                RDVGFDSLTAVELRKRLGAATGLSLPSTMVFDYPTPLELAQYLRAEILGAVLEVAGPV
                ATGGADDEPIAIIGMACRFPGGVSSPEQLWDLVASGTDAISEFPVNRGWQTGHLFDPD
                PDRPGTTYSTQGGFLHEADEFDPTFFGISPREALVMDPQQRLLLETTWESFERAGIRP
                ETLRSTLTGTFVGSSYQEYGLGAGDGTEGHMVTGSSPSVLSGRLSYVFGLEGPAVTVD
                TACSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNPFIAFSRQRALAKDGRCKAFS
                DDADGMTLAEGVGVVLVERLSDAQRNGHPVLAVLRGSAINQDGASNGLTAPNGPSQQR
                VIRQALANARLAPGDIDALEAHGTGTPLGDPIEAQALFATYGRDRDPESALLLGSVKS
                NIGHTQSAAGIASVIKMVMALRHSELPPTLHADAPSSHVDWSAGTVRLLTQARAWPET
                GRPRRAAVSSFGISGTNAHVLLEQAPVADTPAEERPAVAPVPIAAGVVPWVVTARSAA
                ALRGQAERLLAHAETVGTALPAAGPLDIGLSLVSARARFEHRAVVVPPAGTDPLAALR
                AVATDGPSPVVARGVADVEGRTVFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAAL
                AEFTDWSLVDVLRGVVGAPSLERVDVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQG
                EIAAAVVSGALSLRDGARVVALRSQAIGRALAGRGGMMSVALSVDVLEPRLVEFEGRV
                SVAAVNGPRSVVVAGEPEALDALHARLTADDIRARRIAVDYASHSHQVEDLHEELLEV
                LAELAPRTSEVPFFSTVTGDWLDTARMDAGYWFRNLRGRVRFADAVADLLAAEYRAFV
                EVSSHPVLSMAVQEAIDEAGVPAVAAGTLRRDQGGTDRFLLSAAEVFVRGVDVDWAGL
                FEGTGASRIDLPTYAFQHEHLWAVPPAPEAVAAADPDDAAFWTAVEDGDVSALTAALG
                TDEDSVAAVLPALTSWRRARRDRSTVDAWRYRVAWKPLGGTLPHPSLTGTWLLVTADG
                IDDTDVAGALETYGAEVRRLVLDEECVDRAVLRERLAGAEDVTGIVSVLAAAERTDAV
                PGTSLVLGTALTVALIQALGDAEIDAPVWALTRGAVSTGRADELTAPVQAQVTGIGWT
                AALEHPQRWGGTLDLPAALDARAAQRLAAVLSGALGSDDQLAIRPSGVFTRRIVRAEA
```

Query Match 54.1%; Score 133; DB 1; Length 123580;
Best Local Similarity 71.4%; Pred. No. 1.5e-08;
Matches 175; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```
Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||||  ||| |   ||   | ||  || ||| ||| || || ||||||  |    |
Db      29330 CTCGCCGAGCAGCCCGCCGCGGACCACGACGCCATCCTGCTGGCACTGGTCACCGAGAAG 29389

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
                ||| |||||||| |||||| ||  || ||||    ||   |||||||   | ||||||
Db      29390 GCCGCGACCGTCCTCGGCCACGCCGACCACGACCTGCTCGAACCCGACCTGCCCTTCCGC 29449

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||||| |||||||||||| || ||||| ||||| || ||||||| |||| ||       |||
Db      29450 GACCTGGGCTTCGACTCGCTGACCGCGGTCGACCTGCGCAACCAGCTCACCGCGGAAACC 29509

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || |||  ||| ||| |||||||||| ||||||||||||||||||  || ||||| ||  |
Db      29510 GGCCTCACCCTGCCCGCCACCCTCGTCTTCGACCACCCCAACCCGGCCGCCCTCGCCGCC 29569

Qy        241 CACCT 245
              |||||
Db      29570 CACCT 29574
```

```
RESULT 11
AX211739
LOCUS       AX211739               125401 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 35 from Patent WO0159126.
ACCESSION   AX211739
VERSION     AX211739.1  GI:15523950
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 125401)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 35 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .125401
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT    15664 a  49692 c  42871 g  17174 t
ORIGIN
```

Query Match 54.1%; Score 133; DB 6; Length 125401;
Best Local Similarity 71.4%; Pred. No. 1.5e-08;
Matches 175; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```
Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||||  ||| |   ||   | ||  || ||| ||| || || ||||||  |    |
Db      29738 CTCGCCGAGCAGCCCGCCGCGGACCACGACGCCATCCTGCTGGCACTGGTCACCGAGAAG 29797

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
                ||| |||||||| |||||| ||  || ||||    ||   |||||||   | ||||||
Db      29798 GCCGCGACCGTCCTCGGCCACGCCGACCACGACCTGCTCGAACCCGACCTGCCCTTCCGC 29857

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||||| |||||||||||| || ||||| ||||| || ||||||| |||| ||       |||
Db      29858 GACCTGGGCTTCGACTCGCTGACCGCGGTCGACCTGCGCAACCAGCTCACCGCGGAAACC 29917

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || |||  ||| ||| |||||||||| ||||||||||||||||||  || ||||| ||  |
Db      29918 GGCCTCACCCTGCCCGCCACCCTCGTCTTCGACCACCCCAACCCGGCCGCCCTCGCCGCC 29977

Qy        241 CACCT 245
              |||||
Db      29978 CACCT 29982
```

```
RESULT 12
AB070949
LOCUS       AB070949               78210 bp    DNA     linear   BCT 11-OCT-2001
DEFINITION  Streptomyces avermitilis polyene macrolide biosynthetic gene
            cluster.
ACCESSION   AB070949
VERSION     AB070949.1  GI:15824136
KEYWORDS    .
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
```

```
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., StrLm,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Biosynthesis of the polyene antifungal antibiotic nystatin in
            Streptomyces noursei ATCC 11455: analysis of the gene cluster and
            deduction of the biosynthetic pathway
  JOURNAL   Chem. Biol. 7 (6), 395-403 (2000)
  MEDLINE   20334850
   PUBMED   10873841
REFERENCE   2  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
            N-7489, Norway
FEATURES             Location/Qualifiers
     source          1. .123580
                     /organism="Streptomyces noursei"
                     /strain="ATCC 11455"
                     /db_xref="ATCC:11455"
                     /db_xref="taxon:1971"
     gene            complement(46. .783)
                     /gene="nysF"
     CDS             complement(46. .783)
                     /gene="nysF"
                     /function="presumably post-translationally modifies ACP
                     domains on PKS"
                     /note="putative 4'-phosphopantheteine transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysF"
                     /protein_id="AAF71762.1"
                     /db_xref="GI:8050836"
                     /translation="MIELILPATVATEAAYDDRPRPGDRLLSSEREVIARAVESRQRE
                     FTTVRHLARRALRRLGHPDRAILPNRRGAPQWPPGIVGSMTHCAGYRAAAVSPAELSA
                     AVSIDAEPNGPLPAGVLNAIALPSERPHLVALAAHRPDVHWDRLLFSAKESVFKAWYP
                     LTQRELDFSEAEIVIDPTQGAFTARLLVPGPLLGGRRVTVFPGRWHSTPALLTTAVHL
                     PAPTPRRDREHRTHLTVNSPLPRPTFG"
     gene            complement(867. .2684)
                     /gene="nysG"
     CDS             complement(867. .2684)
                     /gene="nysG"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysG"
                     /protein_id="AAF71763.1"
                     /db_xref="GI:8050837"
                     /translation="MASPDDLEEERTAPRPARRLVGLLRPHRRSVALAVSMGVGGIVL
                     NAFGPLLLGRVTDLIADGVLGGVPGPAPGIDFAAIGRLLLVLLALYVVASLFMLAQGR
                     LVASAVWRTIHELRRDAREKLTRLPLRHFDRQPAGELLSRTTNDIDNLQQTLQQTLAE
                     LITSIFSLLTMLVLMLVISPSLAVVMLLSVPVSALIAARISKRAQPHYAAQWSANGTL
                     NAHVEEVCTGHALIKGFDRRAAAEERFDACNDAVYRAAAKAQFASGAMEPVMMFVANL
                     GYVLVAVIGAWKVINGTLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGIASAQRVF
                     TLLDAPEQAPDPLRPGTPARAEGRVEFTDVSFRYSPDTPLIENLSLTVEPGSTVAIVG
                     PTGAGKTTLGNLLMRFYEPDSGRILLDGTDTATMTRDDLRSRFGLVLQDTWLFGGTIA
                     ENIAYGAPGACRADIEEAARATCADRFIRTLPQGYDTVLDDESGTVSAGEKQLLTVAR
                     AFLARPAVLVLDEATSSVDTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIVVM
                     DAGRIVEQGTHDQLLCAQGLYARLHAARTHTPTAGAAAG"
     gene            complement(2662. .4416)
                     /gene="nysH"
     CDS             complement(2662. .4416)
                     /gene="nysH"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysH"
                     /protein_id="AAF71764.1"
                     /db_xref="GI:8050838"
                     /translation="MLLRLLRAQLRPYAWATAALVALQLVQILGTLLLPTLGAALIDQ
                     GVVRGDGGRITELGVVMGVVALVQIAAALGAAALAARTATAMGRDLRSALFRRILDFS
                     AREIGRFGTPSLLTRSVNDVQQVQNLAQTGFGIVVCAPLMCLGSVLLALRQDVPLALL
                     LVALVLVVAVCFGLLLARMGTLYARMQLTLDRLGRLLREAITGVRVVRSFVRDDHERA
                     RFAQTNDAFLVVSRRVGRLIATMLPVVLLLMNGFTVALLWTGSHRIDAGRMPIGSLSA
                     LLSYLSLILMSVVMLAFVFLSVPRARVCAGRIAEVLDTGSSVAPPAAPQPVRGPAGRI
                     ELCAAGYRYPGAEEPVLRDVDLTVEPGERIAVLGSTGSGKTTLLNLVLRLADATEGAV
                     RVGGTDVRELTAATLAAAVGFVPQRPYLFSGTVASNLRFGRPDATDEELWEALRVAQA
                     ADFVARMPDGLDAEITQGGGNVSGGQRQRLSLARALLRRPEIYLFDDCFSALDQATDA
                     ALRTALVPYTAGATVITVAQRISAGRDADRIVVLDRGRVVAQGTHDVLLRTSPTYREI
                     ALSQLTEEEAAHGLAGRP"
     gene            4714. .5748
                     /gene="nysDIII"
     CDS             4714. .5748
                     /gene="nysDIII"
                     /function="presumably involved in mycosamine biosynthesis"
                     /note="putative dGDP-mannose-4,6-dehydratase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysDIII"
                     /protein_id="AAF71765.1"
                     /db_xref="GI:8050839"
                     /translation="MSKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKSRV
                     SRLASELDFIDGDLMDQGSLVSAVDTVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                     VLRMLEAIRMVSGLSTSRTVSPRGQIRFYQASSSEMFGKAAETPQRETTLFHPRSPYG
                     AAKAYGHYITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKQGLQDKL
                     ALGNLDAVRDWGYAGDYVRAMHLMLQQDAGDDYVIGTGQMHSVRDAVRIAFEHVGLNW
                     EDYVVIDPDLVRPAEVEVLCADSAKAQDRLGWKPDVDFPTLMRMMVDSDLAQVSRENQ
                     YGDVLLAANW"
     gene            5930. .34363
                     /gene="nysI"
     CDS             5930. .34363
                     /gene="nysI"
                     /function="responsible for condensation steps 9 to 14 in
                     the nystatin polyketide backbone synthesis"
                     /note="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysI"
                     /protein_id="AAF71766.1"
                     /db_xref="GI:8050840"
                     /translation="MDNEQKLRDYLKLATADLRRTRRRVHKLESAAQEPVAIIGMTCR
                     YPGGVRSPEDLWRMVEAGEHGVTPFPTDRGWDLEALAAAPTASGGFLHDAPDFDADFF
                     GISPREAVAMDPQQRVVLESAWEAFERAGIDPTSVKGSRTGVFIGAMAQDYRVGPADG
                     AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                     AGGVTIMSGPGTFIEMGRQGGLSADGRCRSFGDTADGTGWAEGVGILVLERLSDAVRN
                     GHEILAVVRGTAVNQDGASNGLTAPNGPSQQQVIQQALVNARLAAGDIDVVEAHGTGT
                     TLGDPVEAQALLATYGQNRPADRPLLLGSVKSNLSHTQAAAGVAGVIKMVMAMRHGTL
                     PRTLHAEEPTHHVDWSQGAVRLLTDTTDWPATGAPRRAAVSSFGISGTNAHTIIEQAP
                     EPQPEDAATAQDDAAGSTPATAPVVPGVVPVLLSGRTPDALRGQAAALRAALDTGRRP
                     DLLDLAHSLATTRAGFEHRAVLLATDHPALTDGLTALADADDPAAAPAWITGTTRAET
                     RLAVLFTGQGAQRLGAGRELAARFPAFATALDAALDAFTPHLDRPLREVLWGTDAALL
                     DRTAYAQPALFAVEVALYRLIESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAATLVA
                     ARGRLMQALPDGGAMIAVQASEADVAPLLAGHEDQVAIAAVNGPSAVVLSGAEATVTA
                     LAEQLAADGRKTRRLRVSHAFHSPLMEPMLDAFRAVVEDLTLQPPLLPVVSNLTGKPA
                     TVAQLTSADYWVDHVRHAVRFADGIDWLARHDTTAFLELGPDGVLSAMAQDCLDAADA
                     DAVTLPALRAGRPEEHTLTTALAGLHVHGATLDWTGCFAGTGARRTDLPTYAFQRRRY
                     WPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRG
                     TTLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRT
                     VNVHARPDTGDDTPWTAHATGVLTTADASRQLPASSEQGGTPLAGDPHPALDAAQWPP
                     AGAEPLPLDGHYDRLADGGFGYGPVFQGLRAAWRGGDVVYAEVELPEAGRSDAEAFGL
                     HPALLDAALHAAPFTGLGERGRGGLPFSWEGVSLHAGGATTLRVRLTPVADDALALTV
                     ADGTGAPVLSVDSLVLRSVATQQLDTAAAVARDALFRLDWTPVQPTATDPGPVALLGA
                     DPFGLLTHAGPADAPAYPDLAALAAADGPVPTTVVLSLAGTGDDAADPARSAHRCAAE
                     ALAAVQTWLDHHERFAAARLVFVTRGATVGRDVAAAAVWGLVRSAQSENPGCFALVDL
                     DPDGAVGAAALVAALVSGEPQLAVRGDVLRVARLVRRPLTEVGAGADGTGDGVGDGSG
                     VSFSGEGAVLVTGGTGGLGAVLARHLVAEYGVRDLLLVSRSGERAVGAGELVAELAGV
                     GARVRVVACDVTDRAAVVELVGGHAVSAVVHAAGVLDDGMVGALTGERLSAVLRPKVD
                     AVWHLHEATRGLDLDAFVVFSSLAGVFGSPGQANYAAANAFLDALMTRRRAEGLPGLS
                     LAWGPWEQSGGMTGTLTDVDAERLARSGVPPLSVAQGLALFDAAVAGTDATCVPVRLD
                     LPVLRARGEVPPLLRSLIRVRARRAAVAGSATAGNLAQRLRRLDEDGRDEMVLDLVRG
```

extension modules 9, 10, 11, 12, 13, and 14"
/codon_start=1
/transl_table=11
/product="AmphI"
/protein_id="AAK73501.1"
/db_xref="GI:14794893"
/translation="MDNEQKLRDYLKLATADLRRARRRVGELESASQEPIAIVGMTCR
YPGGVSSPEDLWRMVEAGENGVTPFPTDRGWDLEALASSPTSRGGFLHDAPEFDADFF
GISPREAVAMDPQQRVVLESAWEAFERAGINPTSVKGSRTGVFIGAMAQDYRVGPADG
AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
AGGVTVMSGPGTFIEMGRQGGLSVDGRCRSFGDTADGTGWAEGVGILVLERLSDAIRN
GREILAVVRGTAVNQDGASNGLTAPNGPSQQAVIEQALVNARLSAGDIDVVEAHGTGT
TLGDPVEAQALLATYGQQRDEDKPLLLGSVKSNISHTQAAAGVAGVIKMVMAMRHGVL
PRTLLADEPTRHVDWSQGAVRVLTENTEWPATGAPRRAAVSSFGISGTNAHTIVEQAP
EPEPADPEDDAPSTPAAVTGVLPVLLSGRSPEVLRAQAAALLTTLGTGTPPAPADLAY
SLATTRTAFEHRAVLLASDLPELTGRLTAIAEGTDPAVLADTVTGTARTETRLAVLFT
GQGAQRLGAGRELAARFPVFAAALDAALDAFTPHLDVPLRKVLWGEDADRLDRTEYAQ
PALFAVEVALYRLLESFEVKPDHLAGHSVGEIAAAHVAGVFSLDDAATLVAARGRLMQ
ALPEGGAMVAVQASEDEVAPLLAGHEDLVSLAAVNGPSAVVLSGDETTVTELAARLAA
DGRKTSRLRVSHAFHSPLMAPMLDEFRNVVEGLTLHSPLLPVVSDVTGEPATVAQLTS
PDYWVDHVRQAVRFADGIDWLARHDVTAFLELGPDSVLSAMAQNCLDAAGSDALTVPA
LREGRPEDHTFTAALAALHTQGTALHWDACFTGTGARRTDLPTYTFQRRRYWPRAVQG
GAADLRSVGLGAAHHPLLSAAVSLADSEGALLTGRISLLSHPWLADHTVRGATLLPGT
AFLELAVRAGDEVGCDRVDELTLAAPLVLPEQGGVQVQLWIGNPDASGRRSVTVYGRP
DADEDAPWTSHATGVLSASRTTSDFDATVWPPADAETLPVDGLYERLAEGGFGYGPLF
QGLRAAWRRGDEVFAEVVLPESGHTDAESFGLHPALLDSALHAASFVDLDERAAGGLP
FSWEGVSLHASGATTLRVRLAPAAGDAVAIAVADDSGQLVLSADSLILRAVAAREIDA
AAALVRDALFRLDWVPVTAVAASGTAAALVGEDPFGLRALPQFGDLAVHPDLADLAAA
DGAVPDTVLLPLTGTGPDADPVTAAHRAATEALAAVRTWLEQDERFAASRLALVTRGA
TTGHDPAAAAVWGLVRSAQSENPGRFLLVDLDADQDTPALPAAALTSEEPQLAVRGEE
LRAARLVRRPASTAEAVPAFGGEGAVLVTGGTGGLGAVLARHLVAEHGVRELVLVSRR
GGAAAGAAELVAELAESGARATVVACDVTDRAAVAELVAAHPVSAVVHSAGVLDDGMV
GTLTPERLTTVLRPKVDAAWNLHEATRDLDLKAFVLFSSVAGVLGSPGQANYAAGNAF
LDALAAHRRAAGLPGLSLAWGPWEQTGGMTGGISEDDLRRMARAGTPALTVEQGLALL
DAALDGDDAALAPVRLDLSVLRAQGEVPPLLRSLIRGRSRRAAVAGSATAGGLAQRLA
RLDAESRDELVLDLVRGQVALVLGHATGAEIDAGRAFRELGFDSLTAVELRNRLNTVT
GLRLPATLVFDYPTVSHLASYVLDELLGTEVEAEVVQRGTAAVADDPIVIVGMACRYP
GGVTSPEDLWRLVTEGTDAVSGFPVNRGWDVENLYHPDPDHPGTAYTRSGGFLHEAGE
FDPGFFGMSPREALATDSQQRLLLEASWEAIERAGIDPVGLRGSATGVFAGVMYSDYS
AMLGSPEFEGFQGSGSSPSLASGRVSYTLGLEGPAVTVDTACSSSLVAMHWAMQALRS
GEISLALAGGVTVMSTPAVFVDFARQRGLSPDGRCKAFSDSADGVGWSEGVGMLVLER
QSDAIRNGHQILAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALASGGLSAGDVDVV
EAHGTGTTLGDPIEAQALLATYGRDRDPEQPLLLGSVKSNIGHTQAAAGVAGVIKMVM
SMRHGVLPRTLHVDAPSSHVDWTEGAVELLTEQTAWPETGRPRRAAVSSFGISGTNVH
TVLEQAPGTTVPAPAAPERTAGAVPLLLSGRTRDALRAQAARLLTHLQNHPEPSLADL
GHSLATTRSRFERRAAVIAQDREGLLASLGSLAAGRPDPAVVEGEAAGRARVAVMFSG
QGSQRAAMGRELYETQPRFAAAFDEVCAALDPLLDRPLREVVFAAEGSEEAALLDRTG
WTQPALFAVEVALYRLVESWGVRADFVTGHSIGEIAAAHIAGVFTLQDAARLVAARAT
LMEALPSGGAMVAVQATEEEVAPLLGEGLSVAAVNGPTSVVVSGDEDPAVELAAEFSG
RGRRTKRLRVSHAFHSPHMDAMLDAFRTVAETLSYAAPRIPLVSDLTGRRADDAEVRT
ADYWVRHVREAVRFADCVRTLRDAGATLFLELGPDGLLTAMAEDTLGDERYDHNTALV
PLLRADRPEESAAATAAARLQIHGVDLDWTAYLAGTGARRVDLPTYAFQHAHYWPQLP
SAAPSPAGDPADQKLWAAVERGDAAELAAVLGLDEDSLTPLDSLLPALSSWRRGNQEK
ALLDTLRYRVEWTRLSKPAAPVLDGTWLLVSSDATADDETELLDGLAEALGAHGARVR
RLVLDADCADRAVLGARLADTENADNTAQVLSVLPLDERPTDGPAGFTQGLALTIALV
QALADTGAHGRLWTATRGAVSTGPADPVTHPAQATAWGMGRGVALEHPRLWGGLVDLP
ADFDRGAGQRLAEVLAVKDAPDGEDQVALRATGVHGRRLVRHIVDELPSADQFTASGS
VLITGGTGGLGAETARWLARSGAAHLVLTSRRGPDAPGAAELRAELEQSGASVSIVAC
DVADRDALAAVLDGLSADQPLTGVVHTAGVGHYGPLDALTPAEFAGLTAAKLAGAAHL
DNLLGDRELDFFILFGSIAGVWGSGDQSAYGAANAYLDALALARRARGLAATSIAWGP
WGGTGMAADDAVSGTLRRQGLGLLDPAPALTEMRRAVVRQDVTVTVADVDWTRYAPLF
TSARPSALISDLPEVRALAAENTPADTGDASEIVQRVRSLSEPEQLRLLTDLVRTEAA
TVLGHSSAGAVPEDRAFREIGFDSLTAVELRKHLGAATGLSLPSTMVFDYPTPLELAQ
YLRAEMVGSVLEVAGPVATGGTDDEPIAIIGMSCRYPGGVSSPEQLWDLVLSGTDAIT
DFPVNRGWNTAGLYDPDPDHPGTTYSTQGGFLHEADEFDPMFFGISPREALVMDPQQR
LLLETTWEAFERAGLTPDTLRGSLTGTFIGSSYQEYGMGAGDGAEGHLVTGTSPSVLS
GRLAYVFGLEGPAVTVDTACSSSLVALHLACQALRNGESNLAVAGGATVMTTPNAFVA
FSRQRALAQDGRCKAFSESADGMTLAEGVGIVLVERLSDARRNGHPVLAVIRGSAINQ
DGASNGLSAPNGPSQQRVIRQALANARVAPGEIDLLEAHGTGTPLGDPIEAQALFATY
GRTRTPETALLLGSVKSNIGHSQSAAGVASIIKMVMALRHGVMPQTLHADEPSSHVDW
SPGTVRLLGENTDWPQTGRPRRAAVSSFGISGTNAHVILEQETEAPAAEDEQLAPAPL
PVAAGVVPWLLSARGAAALREQADRLLTHLVTADPAARPIDIGLSLATSRALFEHRAV
VVPPAGTDPLEALRAVAADGPSGVVARGVADVAGRTVFVFPGQGSQWAGMGAQLLDES
PVFAERIAECAAALAEFTDWNLIDVLRGAEGAPTLERVDVVQPASFAVMVSLAAVWRA
QGVEPDAVVGHSQGEIAAAVVSGALSLRDGARVVTLRAQAIGRSLAGRGGMMSVALPV
AEVEARLEAFEGRVSVAAENGPRSSVVAGEPEALDELHAQLTAEEIRARRVAVDYASH
SPHVEDLHDEILELLAEVAPRTSEIPFFSTVTGDWLDTTVMDAGYWYRSLRGRVLFAD
AVRDLIAADHRAFIEVSSHPVLAMSVQDMIDDAGVAGVASGTLRRDNGGLDRFLLSAA
EVFVRGVQVDWAAVFEGTGASRVDLPTYAFQHENLWAMAAAPEAVTAADPEDAAFWTA
VEDGDVSALTAALGTDEDSVAAVLPALSSWRRARKERSTVDSWRYRPTWKPVTKLPQR
TLDGTWLLVSADGVDDTDVAEALETGGAEVRRLVLDESCTDRAVLRERLTDADGLTGI
VSVLAGAERTGAVPGTGLVLGVALTVALVQALGDAGIDTPLWALTRGAVSTGRSDKVT
APVQAQVTGIGWTAALECPGRWGGVVDLPETLDARAGQRLAAVLAGALGDDDQIALRS
SGVFTRRIVRADAAPDGSARDWKPRGTTLVTGGSGTLAPHLARWLAEQGAEHLVLVSR
RGPEAPGAAELRAELAERGTETTLAACDITDRDAVAALLESLKAEGRTVRTVVHTAAT
IELHTLDATTLDDFDRVLAAKVTGAQILDELLDDEELDDFVLYSSTAGMWGSGAHAAY
VAGNAYLAALAEHRRARGLTALSLSWGIWADDLQLGRVDPQMIRRSGLEFMDPQLALS
GLKRALDDDEQVIAVADVDWETYHPVYTSARPTPLFDEVPEVQRLTAAAEQSAGDPAR
GEFAAALLALPAAEQHRKLLETVRTEAASVLGLSSAEDLTDQRAFRDVGFDSLTAVGL
RNRLASVTGLTLPSTMVFDYPNPAALAGFLHSELADVHSAGAVAVTAGAPVDDDPIAI
VGMSCRYPGGITSAEQLWRVSLEEVDAVSVFPADRGWDAEALYDPDPDASGRTYSVQG
GFLRDVADFDPGFFGISPREALSMDPQQRLLLETAWEVFENAGLDPVAQRGSRTGTFI
GASYQDYGAAVPGSEGSHGMITGSLPSVLSGRVSYLFGLEGPAVTLDTACSSSLVAI
HLACQSLRNGESTLALAGGASIMSTPMSFIGFSRQRALAEDGRCKAYAEGADGMTLAE
GVGLILLERLSDARRNGHEVLAVIRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAG
VEANDIDVLEGHGTGTALGDPIEAQALFATYGKDRDPERPVLLGSVKSNIGHTQMASG
VASIIKLVHALREGVAPKSLHIDQPSTHVDWSSGTIQLLTERTEWPETGRPRRAAVSS
FGLSGTNVHTVLEQAPAADAPAAEDTPAPRDALVPVLVSGRGEAALRAQAGALLDLLA
ERPGIHPTDLAFSLATSRAALEHRAAVVADDHEALVRGLTALRDGLPGAGLVQGRTGR
GRTAFLFTGQGSQRLGMGRELYERHPVFADALDAVLARIDGTTERPLRDVLFAAEGSQ
DAALLHRTGYAQPALFALEVALFRLLESWGVTPDYLAGHSVGEIAAAHVAGVLDLDDA
CTLVAARGRLMQALPEGGAMVALEAAEDEVLPHLEGLADQVSVAAVNGPRSVVVAGEE
EPVLALAAHFAEQGRRTKRLRVSHAFHSPLMDPMLDDFAAVARALTYHAPSIPFVSNV
TGTLAAPEQVCTADYWVSHVRSAVRFADGIGWLSTQGGVQTFLELGPDGVLSGMARES
LTDASRTALLPTLRGDRPEEQALVTAVAAAHAHGFDVDWTAWFQGSGARRVALPTYAF
QRERYWPDTTAAGITAPAPGSALDAEFWAAVEHADVASLTASLGLDDATVTAMVPALT
AWRQRRGEQSALDSWRYRVTWKPRGGAPGAAPTGRWLVLVPAEHRDEATAAWAADVEA
ALATATVRVEVTGTDRAALAARLTEAADGDTFQGVLSLLALAPGDAGHPGAPAALTLT
ATALQALGDARIDAPLWNITRGAVAVGRSEQVTAPEQAAVWGLFRAAALELPARVGGS
VDLPEDLDTQAARRLRGILAAADGEDAVAVRASGVFLRRLAHSPAADTVGSAFDPAAG
TVLITGGTGGIGGHLARRLARDGAAHLLLTSRRGPDAPGAGELRAELEESGARVTIAA
CDAADRDALAALLATVPEDAPLTAVFHTAGVVDDHVVDELTPESFATVLHAKTVAARH

Query Match 54.1%; Score 133;. DB 1; Length 113193;
Best Local Similarity 71.4%; Pred. No. 1.5e-08;
Matches 175; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

```
Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||||| |    | ||| || |  |||     ||| |||| ||| |||||  |   |
Db      90141 CTCGCCGGGCTCGACCGCGAGCAGCGCCACGAGGTCCTGCTCGACCTCGTCCGTACGAGC 90200

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
                ||||   ||||| |||||| || ||  |  |  ||||  | |||||||||| ||||||
Db      90201 GCCGCCCTGGTCCTCGGCCACGCCGCCGGCTCCGACATCGACGCCGACCGCGCCTTCCGC 90260

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||||||||||||||||||||||||| |||||||| |||||||| ||   |  | |||||
Db      90261 GACCTCGGCTTCGACTCCCTCACCGCGGTCGAACTCCGCAACCGCCTGCGCACCGCCACC 90320

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || || |||||  |  ||||| | | ||||||||| |||| ||   || | ||| || ||
Db      90321 GGTCTGCGCCTGTCGGCCACCATGGTCTTCGACCATCCCACCCTGTCCGCGCTCGCCGAC 90380

Qy        241 CACCT 245
              |||||
Db      90381 CACCT 90385
```

RESULT 10
AF263912
LOCUS AF263912 123580 bp DNA linear BCT 24-MAY-2000
DEFINITION Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster, complete sequence.
ACCESSION AF263912
VERSION AF263912.1 GI:8050835
KEYWORDS .

```
Qy        241 CACCTC 246
              ||||||
Db      15503 CACCTC 15498


RESULT 8
AX250262
LOCUS       AX250262               30000 bp    DNA     linear   PAT 05-OCT-2001
DEFINITION  Sequence 2 from Patent WO0168867.
ACCESSION   AX250262
VERSION     AX250262.1  GI:15984064
KEYWORDS    .
SOURCE      Streptomyces cinnamonensis.
  ORGANISM  Streptomyces cinnamonensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 30000)
  AUTHORS   Leadlay,P.F., Staunton,J. and Oliynyk,M.C.
  TITLE     Polyketides and their synthesis
  JOURNAL   Patent: WO 0168867-A 2 20-SEP-2001;
            Biotica Technology Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .30000
                     /organism="Streptomyces cinnamonensis"
                     /db_xref="taxon:1900"
                     /note="Nucleotides 30001 - 60000 of the monensin
                     biosynthetic gene cluster"
BASE COUNT     3911 a  10993 c  10862 g   4234 t
ORIGIN

  Query Match             54.5%;  Score 134;  DB 6;  Length 30000;
  Best Local Similarity   71.5%;  Pred. No. 1.6e-08;
  Matches  176;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              || ||| | | | |   ||   | |   ||  | |||||  ||  ||  | |||   ||
Db       5554 CTCGCCCGGCTGTCCGCCGCCGACCGCGACCGCGCCCTCAACGAACTCATTCGCGAGCAG 5613

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
              ||||| ||||||||||  ||| || | ||||| |||||    |  | ||||||  ||||
Db       5614 ATCGCGACCGTCCTGGCACACCCCTCACCCGACACCATCGAACTGGGCCGCGCCTTCCAG 5673

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||  | ||||||||||| ||||||||| | ||||| |||||||| ||||||    |||||
Db       5674 GAGTTGGGCTTCGACTCGCTCACCGCCCTGGAACTCCGCAACCGCCTCTCCACGGCCACC 5733

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              ||  |||| || ||| ||||||||| |||||||||| | |||||| |||||    | |
Db       5734 GGCATCCGGCTGCCCGCCACCCTCGTCTTCGACCACCCGAGCCCCACCGCCCTCGTACGC 5793

Qy        241 CACCTC 246
              || |||
Db       5794 CATCTC 5799


RESULT 9
AF357202
LOCUS       AF357202               113193 bp    DNA     linear   BCT 17-JUL-2001
DEFINITION  Streptomyces nodosus amphotericin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF357202
VERSION     AF357202.1  GI:14794889
KEYWORDS    .
SOURCE      Streptomyces nodosus.
  ORGANISM  Streptomyces nodosus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 113193)
  AUTHORS   Caffrey,P., Lynch,S.V., Flood,E.M., Finnan,S.M. and Oliynyk,M.
  TITLE     The amphotericin biosynthetic gene cluster from Streptomyces
            nodosus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 113193)
  AUTHORS   Caffrey,P., Lynch,S.V., Flood,E.M., Finnan,S.M. and Oliynyk,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2001) Industrial Microbiology, University College
            Dublin, Belfield, Dublin, Ireland
FEATURES             Location/Qualifiers
     source          1. .113193
                     /organism="Streptomyces nodosus"
                     /db_xref="taxon:40318"
     gene            complement(4. .1824)
                     /gene="amphG"
     CDS             complement(4. .1824)
                     /gene="amphG"
                     /note="possible amphotericin export protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphG"
                     /protein_id="AAK73498.1"
                     /db_xref="GI:14794890"
                     /translation="MAPDVPEEHEEERESEQPVRRLAALLRPHRRSVGLALTAGVVGI
                     LLNAFGPLLLGRVTDLIADGVLGHGGPAPGVDFGALGRLLMILLVLYVVASVFMLVQN
                     WLVASVVRLLIHDLRHRAQEKLARLPLRYFDRKPAGETLSRGTDDVDNLQQTLQQTLT
                     DLISSVFSLVIMLSLMLIISPSLAGVMLLSVPVSGLLAAWISKRAQPQYAAQWSASGK
                     LTAHVEEMCAGHALVKAFDRRAEAEQRFDERNEAVYRAGSGAQFASGAIEPVMMFVAN
                     LGYVAVAVVGAWKVVNGSLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGVASAQRV
                     FTLLDAPEQEPEPDRPLAVERVEGRVEFQDVSFRYSPDTPLIEGLSLSVEPGSTVAVV
                     GPSGAGKTTVANLLMRFYEIDSGRILLDGTDTAAMNRDDLRSRFGLVLQDTWLFKGTI
                     AENIAYGSPGATRADIVEAARATYADRFIRTLSQGYDTVLDDESGGVSAGEKQLITVA
                     RAFLARPAVLVLDEATSSVDTRTELLIQRAMNTLRPGRTSFVIAHRLSTIRDADVIVV
                     MKSGRIVEQGTHDQLIDAQGAYARLHAARADAPAADVTVG"
     gene            complement(1805. .3628)
                     /gene="amphH"
     CDS             complement(1805. .3628)
                     /gene="amphH"
                     /note="possible amphotericin export protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphH"
                     /protein_id="AAK73499.1"
                     /db_xref="GI:14794891"
                     /translation="MAPSVRWPTPFQRAAAGRRQPGCHSVLLRLMRTQLRPYAGSVVA
                     LVVLHLVQILGTLLLPTLGAALIDEGVVRHDSDRIGTIGATMAVVALVQIAAALGAAA
                     LGARTSTALGRDLRSAVFRRVLDFSAREIGRFGTPSLLTRSVNDVQQVQNLAQSGLGI
                     FVAAPLMCLGSVLLALRQDVTLALILVPMVLVVAVCFGLLLSRMAALYARLQQTLDRI
                     GRLLRERITGVRVVRSFARDAHEGERFTRTNEELLGLSLGVGRLIAVMLPSVLLLMNL
                     FTLGLLWVGARRIDSGSMQIGALSAFLSYLSLILMSVVMLAFVFLNVPRARVCAERIT
                     EVLQAETDVVPPASPRPMAGPAGQVELVGAEFRYPGAENAVLRDLSLTLRPGERVAVL
                     GSTGSGKTTLLHLILRLVDVTAGEVRIGGTDVRELDPSVMAAAVGYVPQRPYLFAGTV
                     ASNMRFGRPDATDEELWEVLRIAQHEGFVTRLGGLDTEIAQGGTTVSGGQRQRLAIAR
                     ALLRRPAIYLFDDSFSALDQSTEAALRKALVPYTEGATVITVAQRVASVRDADRIVLL
                     DQGGIAATGTHDALLRDSPTYREIALSQRTREETAHGAGRS"
     gene            3840. .4874
                     /gene="amphDIII"
     CDS             3840. .4874
                     /gene="amphDIII"
                     /note="mycosamine biosynthesis protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphDIII"
                     /protein_id="AAK73500.1"
                     /db_xref="GI:14794892"
                     /translation="MPKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKFRV
                     SRLASELSFVDGDLMDQGSLVSAVDKVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                     VLRVLEAIRMVSGLSMSRTAGTEGQIRFYQASSSEMFGKVAETPQRETTLFHPRSPYG
                     AAKAYGHFITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKLGLQDKL
                     ALGNMDAVRDWGYAGDYVRAMHLMLQQDAPDDYVIGTGEMHTVRDAVRFAFEHVGLDW
                     KDYVVVDPDLVRPAEVEVLCADSSKAQAQLGWKPSVDFQELMRMMVDADLASVSRQNE
                     LDDLLLAHSW"
     gene            5042. .33574
                     /gene="amphI"
     CDS             5042. .33574
                     /gene="amphI"
                     /note="polyketide synthase multienzyme polypeptide housing
```

```
                     /gene="pimS4"
                     /function="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="PimS4 protein"
                     /protein_id="CAC20919.1"
                     /db_xref="GI:12055070"
                     /translation="MPNDKPNDKKLVDYLKWVTNDLHQTRQRLREVESAKQEPIAIVG
                     MSCRLPGGVRSPEEFWQLLKDGRDGISTFPTDRGWDLDLLAGDGSGSSATAEGGFIDA
                     ASFDAGFFGISPREAVTMDPQQRILLESSWEALERAGIDPVSLRGTETGVFVGTSGID
                     YINTVMNSREDIEGHGSTGLIASILSGRVSYTFGLEGPAVTVDTACSSSLVSLHLAAQ
                     ALRVGECSLALAGGVTAISTPMSFVGFSRQGGLAPDGRCKAFSDTADGTAWSEGVGVL
                     VVERLSDARRNGHPVLAVVRGSAVNQDGASNGLTAPSGPSQQRVIRQALAAAGLSASD
                     VDVVEAHGTGTALGDPIEAQALLATYGRDRDPERPLLLGSVKSNLGHTQAAAGAAGLI
                     KMILAMRHGVLPRTLHVEAPSSHVDWSSGALQLLTEPQEWPKADRPWRAGVSSFAISG
                     TNAHIILEQPPAAEEPEDEEGPHTSPSAVPWLVSAKSEEALTAQLDRISALTDRSPLD
                     LGLSLATTRAHLEHRAVLLSEGGETVEVARGHAEHTAGKLAVLFSGQGAQRVGMGREL
                     YGRFPVFAEALDGVLAHFDAGLRDVMFGEAEGLDETGFTQPALFAIEVALFRLSESLG
                     VRPDFVAGHSIGEIAAAHVAGVFSLADACALVAARARLMQALPAGGAMVAVQAAESEV
                     AARLVDGVSIAAVNGPEAVVIAGEEREVLRIAEEFTGLGRKTRRLPVSHAFHSPLMEP
                     MLEDFRRVAEGLSYEAPRIALVSNLTGELATEDLVRSAEYWVRHVRETVRFADGVSAL
                     AAEGASVFLELGPDGVLTAMAQHTLDGTATTVPALRKDRSEETALLTALAQLHVAGVG
                     VDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVVGAGLRSAEHPLLGATLTLA
                     ESGEVVLTGRLSLQAYPWLTDHTMGGSVLFPATGFLELAMRAADQVGCDRVDEFVLMT
                     PLVLAEESGTQVQVLVGPADEAGTRTISIHARPDGDPDQPWMQHAGGTLTSGERIVDF
                     GSTVWPPQGAEAVDLEGFYDTTGYGPAFQGLRSVWQCEDGAYVEVALPDQVAEDAGAF
                     GLHPVLLDAVLQSRRMAGIGAAGDTFLPFAWRGVSLHAGGASVLRARVIKTGEDTLSL
                     VAVDAEGAPVLSVEELVLRTQLTHTDAPLTARRGAERDALLSLEWVSGPEVPPAGVRC
                     VSLGADELGVGASVATLADLSGDEDFVVVPVSGVGGDVPAAAHELTARTLGLLQEWSG
                     REGTSASRLVFVTRNAVSADGGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADSG
                     DLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGGT
                     GGLGRELARHLVVERGVRHLLLVSRSGPEASGVGELRDELAAHGAEIAVRACDVADRA
                     AVEAVVGAVDAAHPLTAVIHTAGVLDDGVIASLTPERLSAVLRPKVDAAWNLHEATKD
                     LDLAAFVLYSSISGVLGSLGQGNYAAGNVFLDALAGHRAGLGLPAQSLAWGAWTQGGG
                     MTATLSDADMQRMASYGTAPLSPERGLALFDAATAGSATHLVAVGRVSGPVRAQGPVP
                     SMLRGLVKGARRTAVNTTGETAVSLADRLRAARPEERTQLMTDVVRTEAAAVLGHASS
                     QNVDARREFYELGFDSLTSVELRNRLSTVTGLRLSATVVFDSKNPTDLAARLYADLAA
                     QSALDGSASQAGFLPVAAPETDSLERMFLDALDSGKVPEAQRMLSALGALRPSFENTA
                     ELEDLPLPATLAEGPGTPRLICVSTPTANGGVHEYARLAASFRGERHVSALPLVGFAT
                     GERLPATAETAVRVIAESALRASDGNPFVLVGHSSAGAFAYLAAALLENTWGIRPEAV
                     VLLDTLSLRHEQNESIDYAGLMRRHFMVDEVSPVRMTNSRLSAMARWMGMLNQLEVRH
                     TTVPVLIIRATKETFGIGTDTGIYGEDHGSPVDVRSVDADHFSMVRDDAPETARIVRE
                     WLDSLGNA"
gene                 complement(9772. .15198)
                     /gene="pimS3"
CDS                  complement(9772. .15198)
                     /gene="pimS3"
                     /function="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="PimS3 protein"
                     /protein_id="CAC20920.1"
                     /db_xref="GI:12055071"
                     /translation="MTTSSETVVAALRASLKEAENLRRQNRKLVAAATEPIAIVAMAC
                     RYPGGVQSPEDLWELLTAGKDAIGAFPTDRGWDLDALRDAGVDRRGNAVSQEGGFLDG
                     VGDFDADFFGISPRESVTMDPQQRLLLETSWEAIERAGIDPTTLRGSRTGVFIGTNGQ
                     DYAYLLVRSLADATGDIGTGIAASATSGRLSYTLGLEGPAVTVDTACSSSLVALHTAS
                     HALRAGECSLALVGGVNVMSSPGSLMEFSRQGGLAADGRCKAFSDTADGTGWSEGVGV
                     LVVERLSDARRNGHPVLAVVRGSAVNQDGASNGFTAPSGPSQQRVIRQALAAAGLSAS
                     DVDVVEAHGTGTPLGDPIEAQALLATYGQDRDPERPLLLGSVKSNLGHTQAAAGAAGL
                     IKMILAMRHGVLPKTLYADVPSSHVDWTSGEVELLTEPQEWPKADRPWRAGVSSFGIS
                     GTNAHVIVEQAEPDEEPVAEPGLAPAVVPWLVSAKTEEALAAQLARISSRRGTASPVD
                     VGFSLASGRSAFEHRAVLLAQCPDGELSEAARGRAETSRPLAVLFSGQGAQRVGMGRE
                     LYVRFPVFAEALDEVLAHFDAGLRDVMFGDAEGLDETGFTQPALFAIEVALFRLAESL
                     GVRPDFVGGHSIGEIAAAHVAGVFSLTDACALVAARARLMQALPAGGAMVAVQASESE
                     VAARLVDGVSIAAVNGPEAVVLAGDETAVLRIAEELAAGGRKTQRLPVSHAFHSPLME
                     PMLEDFRRVAEGLSYGAPQIPLVSNVTGELAVEELVCSPDYWVRHVRETVRFADGVSA
                     LEAEGASAFLELGPDGVLTAMARHTLDETKVAVPVLRKDRPEETALLTALAHLHVAGI
                     GVDWSGIFAGTGARRVDLPTYPFQHERYWPRPAVLSGDVAGAGLRSAEHPLLGATLTL
                     AESGEVVLTGRLSLQTHPWLTDHAIGGNVFFPGTGFLELAMHAADQAGCDRVEDLVLM
                     TPLVLAEENGTQVQVLVGPADEAGTRTISIHARPEDDPDQPWMQHAGGTLTSGERVVD
                     FGSTVWPPQDAVAVDLEGFYEGTGYGPAFQGLRSVWSCGDGAYAEVALPDHVAEDAGA
                     FGLHPALLDAVLQSHRIAGVGSADDQFLPFAWRGVSLHAGGASVLRVRVVKTGDDSMS
                     IAAVDAEGAPVLSAEAMVMRAYRDPGTGQAAARHGAERDELLSLEWVSGPEVPPAGVR
                     CVSLGADELGVGASVASLADLSGDEDFVVLPVSGVGGDVPAAAHELTARTLGLLQEWS
                     GREGSSGSRLVFVTRNAVSADEGESVRDVAAAAVWGLVRSAQSESPGRIALVDLEADS
                     GDLAEVLAGLPGLLAAGDAQFVVREGAVRVGRLVESNEGAGVSPREWDADGTVLVTGG
                     TGGLGRELARHLVVERGVRHLLLVSRSGPEAPGVSGLSDELAAHGAEVTVRACDVADR
                     AAVHALVAAVPAEHPLTAVIHTAGVLDDGLVTSLTPERLSTVLRPKVDAAWNLHEATK
                     DLDLAAFVLYSSISGVLGAPGQANYAAGNVFLDALAWHRMGQGLPAQSLAWGAWTQDG
                     GMTATLSDTDMQRMASYGTAPLSRQRGLALFDAACARSAAHLVAVGRVGGPVRASAPV
                     PSVLRGLVKGARRTAVNTTGETAVSLADRLRAARPEERPRLMTDLVRTEAAAVLGHAS
                     AKNVDARREFNELGFDSLTSVELRNRLSALTGLRLSAALVFDYPNPTALADHFVSELV
                     EEAATGPALLAELERLDSALAAGAPDARTRAAVTARLAQILDDWRDAPAEEDGAEVAE
                     RLEAASTDEIFAFIDNELGRHSDR"
gene                 complement(15210. .43733)
                     /gene="pimS2"
CDS                  complement(15210. .43733)
                     /gene="pimS2"
                     /function="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="PimS2 protein"
                     /protein_id="CAC20921.1"
                     /db_xref="GI:12055072"
                     /translation="MDQEQKLRDYLKRASADLRRSRQRVGELEAAAREPIAIVGMSCR
                     FPGGVRSPEELWSLLAADGDGITKSPADRGWELLAEDAGEFSGGFLHDAAEFDASFFG
                     ISPREALAMDPQQRLLLESSWEAFERAGIDPAAVRGTRTGMFVGAMPQEYRVGPDDDV
                     QGFALTGTTTSVISGRLAYFFGAVGPAVTVDTACSSSLVALHLAAHSLRQGECSLALA
                     AGVTVMSSPTTFVEFNRQGGLSADGRCRSFADSADGTGWSEGVGVLVLERLSEARRNG
                     HEILAVIRGSAVNQDGASNGLTAPNGPSQRRVIEQALVSARLSADEVDVVEAHGTGTT
                     LGDPVEAQALLATYGQGRDDDRPLLLGSVKSNLGHTQAAAGMAGVIKMVLAMRHGVLP
                     RTLHVDAPSSHVDWTQGAVRLLTEHVPWPQGEQPRRAGVSSFGLSGTNAHTILEEAPA
                     APSPEAGQEDTAPVAEPLAAGAVPWLVSGRTREALRAQAARLLEHLTSRPGLAPLDVA
                     YSLATARSGLEHRAAFTAAGPDGARAALTALADGTPAPGLVQDTARTRSKLAFLFAGQ
                     GSQRPGMGRELAARFPVFATALDEVLGHFDEGLERPLKDVLFAAEGTPEAALLDQTGY
                     AQPALFAIEVALYRLAESFGITPDFLAGHSIGEIAAAHVAGVFTLADAAALVLARGRL
                     MQALPEGGAMVSLEATEEEVLPLLEDRPDRLSIAAVNGPAAVVIAGEADDVTAVAEHF
                     DALGRRTKRLRVSHAFHSPLMEPMLAEFRTAVSRLTPQAPVIPVVSGLTGTLATVEQL
                     TSPDYWLDHARHAVRFADALRLARRARRRHLPRMGPDGVLSAMAQSCCRGHGHRRRRR
                     LRHARRVRPGRPEAETLTAALAGLHTRGVAVQWEPYFQGTGARRADLPTYAFQRRRYW
                     PKSLPATGGDVRAAGLGAAHHPLLTAAVSVANSDGLLLTGRLSRRTHPWLADHAVRGT
                     VLLPGTAFLELAVRAGDEAGCGRVEELTLAAPLVLPEEGGVQVQVWVGSPDASGRRAV
                     SVHSRPDGPEELPWTQHAAGTLVAGEHHAGFDATVWPPADATPLDLDGFYARMADTGF
                     GYGPLFQGLRAAWRSGDDVYAEVALPDSDGAASTESSAAFGLHPALLDAALHAAAFVD
                     LGEDARGGLPFSWQDVTLHASGASTVRVKLTPDGDDAVALAVADTTGAPVASIGSLVL
                     RTLPDEQIDAAHSLVRDALFELRWSEARQTAHEAPASAAVLGPDLFGLERALAPTDLT
                     VTAVPGPDAAPHAPDVVLAPVAGTTEDDVVTSVHARTAHVLERLRTWLTDERYAASRL
                     VFVTRGALATDARQASDPVAAAVWGLVRAAQSEHPGRFGLLDLDPGTTEPATEPLLRA
                     LALSSDEPQTAVRGTAVLTARLARAQAPQAATEWDPEGTVVLTGATGGLGRVLARHLV
                     AERGVRHLLLVSRRGADADGAGELMAELAAHGAQVTLAACDLADRAAVRRLLSGVPAE
                     HPVTAVVHSAGSLDDGVIASLTPERISAVLRPKADAVWHLHEATRDLDLAAFVVFSSL
                     SGTVGAAGQGNYAAANAFLDALAQLRRAQGLPGLSLAWGPWAPTAGMTGGLTDDDLDR
                     LARMGTPALTEEQGIALFDAATATDAAVLLPTRIDLSVLRVQSEIPPLWRGLIRTPAR
                     RSVVSASETAVTLVQQLSRLAETDRREVVLDLVAGQVAAVLGHAGTADIDPRRPLREL
                     GFDSLTAVELRNRLSAATGLRTAATVIFDHPTVDALAAHLLDGLMGSEAAAAADAPRT

  Query Match          56.4%;  Score 138.8;  DB 1;  Length 84985;
  Best Local Similarity  72.8%;  Pred. No. 3.1e-09;
  Matches  179;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||| | |||| | | | ||   | |    |    || | |  | ||||||| ||| |||
Db      15743 CTGACGGGCCTGCCGCCCGCCAAGCGCTCCGAGGCCGTACGGGACCTGGTGCGCACCCGG 15684

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
                |||  |||| ||||| ||| ||||  ||||| || |   | |||||||  ||||||||
Db      15683 GCCGCGGCCGTACTGGGACACCCCACAGCCGACGCCGTAGGCGCCGACCGGGCGTTCCGT 15624

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              |||||||||||||||||||| |||||||||| ||||||| ||||  |  | | |||
Db      15623 GACCTCGGCTTCGACTCCCTCAGCGCCGTCGAACTGCGCAACCAGCTCGGCGCCGCTACC 15564

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              |||||   |||  || |||| |||| ||||||||||||||||||| || ||||  |  ||
Db      15563 GGACTGAACCTGACCGCCACGCTCGTCTTCGACCACCCCAACCCCGCCGCCCTGGCAGAC 15504
```

```
BASE COUNT     1884 a   4561 c   4005 g   1931 t
ORIGIN

  Query Match           61.8%;  Score 152;  DB 6;  Length 12381;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-10;
  Matches  152;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12230 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 12289

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12290 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 12349

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGA 152
              ||||||||||||||||||||||||||||||||
Db      12350 GACCTCGGCTTCGACTCCCTCACCGCCGTCGA 12381


RESULT 6
AX250263
LOCUS       AX250263               30000 bp    DNA     linear   PAT 05-OCT-2001
DEFINITION  Sequence 3 from Patent WO0168867.
ACCESSION   AX250263
VERSION     AX250263.1  GI:15984065
KEYWORDS    .
SOURCE      Streptomyces cinnamonensis.
  ORGANISM  Streptomyces cinnamonensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 30000)
  AUTHORS   Leadlay,P.F., Staunton,J. and Oliynyk,M.C.
  TITLE     Polyketides and their synthesis
  JOURNAL   Patent: WO 0168867-A 3 20-SEP-2001;
            Biotica Technology Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .30000
                     /organism="Streptomyces cinnamonensis"
                     /db_xref="taxon:1900"
                     /note="Nucleotides 60001 - 90000 of the monensin
                     biosynthetic gene cluster"
BASE COUNT     4041 a  10632 c  11060 g   4267 t
ORIGIN

  Query Match           56.7%;  Score 139.4;  DB 6;  Length 30000;
  Best Local Similarity   73.1%;  Pred. No. 3.5e-09;
  Matches  179;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||  ||| || ||||  |||  |||||||| |||||||||| |||||| ||  | |||
Db       6404 CTCTCCGTCCTGACAGCCGAGGAACAGCACCGCACCCTCCTCGACCTGGTACGGACGCAC 6463

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
                 ||  ||||||| || ||| |   | ||||| || | | | |||||  ||| ||||
Db       6464 GCGGCAGCCGTCCTCGGGCACGCGGGCACCGACGCCGTACGCGCCGACGCCGCCTTCCAG 6523

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              || |||||||||||||||||||||| |||||||| |||||||| ||||||  | |||||
Db       6524 GATCTCGGCTTCGACTCCCTCACCGCGGTCGAACTGCGCAACCGCCTCTCCGCCTCCACC 6583

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              || || ||||| ||| ||||  ||  ||||   ||||| | |||  |  || || || ||
Db       6584 GGCCTGCGCCTGCCCGCCACGTTCATCTTCCGGCACCCGACCCCGTCGGCCATCGCCGAC 6643

Qy        241 CACCT 245
              | ||
Db       6644 GAACT 6648


RESULT 7
SNA278573/c
LOCUS       SNA278573              84985 bp    DNA     linear   BCT 17-MAY-2001
DEFINITION  Streptomyces natalensis pimaricin biosynthetic gene cluster.
ACCESSION   AJ278573
VERSION     AJ278573.1  GI:12055067
KEYWORDS    ABC transporter; cholesterol oxidase; cytochrome P450 monoxygenase;
            ferredoxin; glucosyl transferase; macrolide efflux pump; mycosamine
            dehyralase; orfX; pimA gene; pimB gene; pimC gene; pimD gene; pimE
            gene; pimF gene; pimG gene; pimH gene; pimI gene; pimJ gene; pimK
            gene; pimS0 gene; pimS1 gene; pimS2 gene; pimS3 gene; pimS4 gene;
            polyketide synthase; sensory transduction protein; thioesterase.
SOURCE      Streptomyces natalensis.
  ORGANISM  Streptomyces natalensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 84985)
  AUTHORS   Aparicio,J.F., Fouces,R., Mendes,M.V., Olivera,N. and Martin,J.F.
  TITLE     A complex multienzyme system encoded by five polyketide synthase
            genes is involved in the biosynthesis of the 26-membered polyene
            macrolide pimaricin in Streptomyces natalensis
  JOURNAL   Chem. Biol. 7 (11), 895-905 (2000)
  MEDLINE   20547809
   PUBMED   11094342
REFERENCE   2  (bases 1 to 84985)
  AUTHORS   Aparicio,J.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2000) Aparicio J.F., Molecular Genetics,
            Institute Of Biotechnology - Inbiotec, AV. REAL 1 - LEON, 24006,
            SPAIN
FEATURES             Location/Qualifiers
     source          1. .84985
                     /organism="Streptomyces natalensis"
                     /db_xref="taxon:68242"
     CDS             <1. .2086
                     /note="ORFX"
                     /codon_start=2
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="CAC20917.1"
                     /db_xref="GI:12055068"
                     /translation="TRPDAFREATEPYGALLAPLMPERSADRVEGTKWTSDWSQARFL
                     AHDAVCEVLLTLAGQHPLVLLLEDLQWADTASLDLLRLLSPRCQGHPLGIVLTARKFE
                     TESDATLRRIVSDVLRGPRTATLRLCGLPRQAVGALVEAHAGSGVDPEVVEVLHRRSE
                     GNPYFVMQLLSLVGDARRLHRPDAVDALLAHIPTGVREALNQRFAALPEPVLRVLRLC
                     AVIGAEVDTDLLRRTATCDASVNAGLESAIRAGLLGEDPHHPGQLHFAHALVQETLFD
                     ELPRADRRRLHAKVADALYARRRGHVGDEEIERVAHHSWHAKSALPAAEVLPRLLLAA
                     EHAEQAMAYEQVEIWLRRAVHLVGFLPPDDPATVRLEQKLHIQLGQVLATTRGYGHPE
                     AETALTRGRALSTVTHAPEDPSVLWALCAASLVTGRYDASRQLSGLLRDLSWQTREPV
                     AVLGAAYGEGIVLHVRGMLPQALAELEHGVERADHFANEGHALARTFQHDPRVSCRSY
                     NTFTHWLLGHRQIADARRRELLSLTEYESRPSDRSFALYVDGVVAAWEGDTRTALASS
                     DEGVRVAGDHGLLYWKAMLGVVKGWGLVHAGDRDEGLALMHTSLAELGQSRTYLRHPL
                     HLGLLGQAQHHNGQIEEARTTFRMLLAAVEQRRERVYLHPALPATPLLHELLGHGADE
                     ALADRDPGRGQV"
     gene            2190. .3566
                     /gene="pimK"
     CDS             2190. .3566
                     /gene="pimK"
                     /function="putative glucosyl transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="PimK protein"
                     /protein_id="CAC20918.1"
                     /db_xref="GI:12055069"
                     /translation="MESARRPILFVSLPESGLLNPLLVLAGELSRQGVEDLWFATDEP
                     RRNDVKRIAEGSPVEPASLGEVDSEMSAVTWSDEVYREVTQPSRFKAHRAVVRHTYRP
                     GLQAEKFRRLQAVIDEVQPALMVIDCISGFAVDAAIARNIPYVLSVPFLPSNVLTAHT
                     HFAKSYTPRGFPVPHTGLSRRMTLAQRVANELFKLRTFAMFLNPRLGKVLAEDNRRRN
                     ELGLPKASFMARIEHADLVLCNSLAELDYPFDIPEKMRLVGAMVPPLPEAPDDQDLSR
                     WLDAQSSVVYVGLGTITRLTREQVGSMVEVARRLEDRHQVLWKLPSEQQHLLPPRESL
                     PGNLRVESWVPSQMDVLAHPHVKVFFTHGGGNGFNEGMYFGKPLVVRPLWVDCYDQAV
                     RGQDFGLSLTLDRPQTIDVNDVVDKLTRVLGTPSFYEKAERRAALMRSAGGRETAAGL
                     VLSLPALA"
     gene            complement(3662. .9736)
                     /gene="pimS4"
     CDS             complement(3662. .9736)
```

```
                VSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
                AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
                LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
                RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
                RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
                TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
                AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
                SLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLGLS"
gene            12071. .30790
                /gene="aveA2"
CDS             12071. .30790
                /gene="aveA2"
                /note="contains Ave module 3, module 4, module 5 and
                module 6
                multifunctional polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="type I polyketide synthase AVES 2"
                /protein_id="BAA84475.1"
                /db_xref="GI:5902892"
                /translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
                CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
                DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGINP
                EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
                ACQALRSGECTMALAGGASVMATPFVFTEFSRQRGLAADGRCKAFSAAADGTGWSEGV
                GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
                PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
                GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
                GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
                KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
                ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
                NDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALHRL
                LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
                TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
                FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
                TQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
                TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
                PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
                NRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLTAVPADEPIAIV
                GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
                FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
                LTYHDYAARFPTAPAGFEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
                HLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSAAADGTGWGE
                GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAD
                LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
                VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
                SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAADDVPGEAAGDDAGTGGEATGPA
                AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
                IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH
                GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
                PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
                QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
                QGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
                TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
                PQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
                SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
                QHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
                PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
                LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
                IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG
                TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
                TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
                LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
                GTWQGNGLATGQVSEHLRRRGMFAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
                LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA
                AVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ
                FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI
                GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
                RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
                SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
                TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
                AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
                LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
                HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
                DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
                ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
                GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAAL
                LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
                ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
                HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
                QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
                TLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
                PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
                RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
                ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
                TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
                THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
                TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
                IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
                KATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
                PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD
                IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
                TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
                PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
                ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
                YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
                AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL

  Query Match             99.3%;  Score 244.4;  DB 1;  Length 64957;
  Best Local Similarity   99.6%;  Pred. No. 3e-22;
  Matches  245;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44769 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 44710

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 120
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db      44709 ATCGCCACCGTCCTCGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 44650

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44649 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 44590

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44589 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 44530

Qy        241 CACCTC 246
              ||||||
Db      44529 CACCTC 44524


RESULT 5
AX006889
LOCUS       AX006889               12381 bp    DNA     linear   PAT 06-SEP-2000
DEFINITION  Sequence 1 from Patent WO0001827.
ACCESSION   AX006889
VERSION     AX006889.1  GI:9994904
KEYWORDS    .
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 12381)
  AUTHORS   Kellenberger,J.L., Leadlay,P.F., Staunton,J., Mcarthur,H.A. and
            Stutzman-Engwall,K.J.
  TITLE     Polyketides, their preparation, and materials for use therein
  JOURNAL   Patent: WO 0001827-A 1 13-JAN-2000;
            KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
            STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PFIZER (US); MCARTHUR
            HAMISH ALASTAIR IRVIN (US); STUTZMAN ENGWALL KIM JONELLE (US)
FEATURES             Location/Qualifiers
     source          1. .12381
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
```

```
            PF   24-FEB-1999 JP 1999046961
            PR
            PI   SATOSHI OMURA,HARUO IKEDA
            PC   C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC   (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(14643)
            FT   CDS             (14824). .(31419).
FEATURES             Location/Qualifiers
     source          1. .31422
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     4136 a  10237 c  11677 g   5372 t
ORIGIN

  Query Match             99.3%;  Score 244.4;  DB 6;  Length 31422;
  Best Local Similarity   99.6%;  Pred. No. 3.7e-22;
  Matches  245;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20089 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 20148

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20149 ATCGCCACCGTCCTCGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 20208

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20209 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 20268

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      20269 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 20328

Qy        241 CACCTC 246
              ||||||
Db      20329 CACCTC 20334


RESULT 4
AB032367/c
LOCUS       AB032367               64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
            Sciences, Kitasato University, Microbial chemistry; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .64957
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            101. .12019
                     /gene="aveA1"
     CDS             101. .12019
                     /gene="aveA1"
                     /note="contains a loading module, Ave module 1 and module
                     2
                     multifunctional polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="type I polyketide synthase AVES 1"
                     /protein_id="BAA84474.1"
                     /db_xref="GI:5902891"
                     /translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
                     ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
                     AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
                     VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTR
                     MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
                     MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
                     RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
                     DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
                     LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
                     GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
                     MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
                     GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
                     GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
                     RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
                     GDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
                     LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
                     AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAG
                     LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
                     EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
                     VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
                     AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
                     AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
                     ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
                     PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
                     THHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
                     EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
                     WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
                     LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
                     PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
                     LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGAL
                     ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
                     QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
                     TAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
                     DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSA
                     ARERERPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
                     FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
                     AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
                     YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
                     GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
                     GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
                     GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
                     NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
                     WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
                     TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
                     SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
                     ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
                     YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
                     VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
                     MTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
                     LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
                     VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
                     NLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
                     TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATPLELALHAGTY
                     VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGASGG
                     VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
                     RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
                     EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
                     RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
                     LRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPE
                     DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
                     ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
```

```
                     RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
                     LRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPE
                     DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
                     ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
                     VSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
                     AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
                     LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
                     RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
                     RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
                     TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
                     AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
                     SLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLGLS"
gene                 12071. .30790
                     /gene="aveA2"
CDS                  12071. .30790
                     /gene="aveA2"
                     /note="contains Ave module 3, module 4, module 5 and
                     module 6
                     multifunctional polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="type I polyketide synthase AVES 2"
                     /protein_id="BAA84475.1"
                     /db_xref="GI:5902892"
                     /translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
                     CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
                     DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGINP
                     EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
                     ACQALRSGECTMALAGGASVMATPFVFTEFSRQRGLAADGRCKAFSAAADGTGWSEGV
                     GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
                     PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
                     GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
                     GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
                     KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
                     ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
                     NDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALHRL
                     LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
                     TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
                     FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
                     TQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
                     TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
                     PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
                     NRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLTAVPADEPIAIV
                     GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
                     FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
                     LTYHDYAARFPTAPAGFEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
                     HLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSAAADGTGWGE
                     GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAD
                     LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
                     VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
                     SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAADDVPGEAAGDDAGTGGEATGPA
                     AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
                     IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH
                     GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
                     PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
                     QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
                     QGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
                     TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
                     PQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
                     SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
                     QHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
                     PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
                     LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
                     IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG
                     TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
                     TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
                     LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
                     GTWQGNGLATGQVSEHLRRRGMFAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
                     LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA
                     AVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ
                     FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI
                     GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
                     RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
                     SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
                     TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
                     AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
                     LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
                     HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
                     DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
                     ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
                     GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAAL
                     LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
                     ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
                     HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
                     QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
                     TLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
                     PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
                     RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
                     ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
                     TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
                     THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
                     TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
                     IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
                     KATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
                     PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD
                     IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
                     TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
                     PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
                     ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
                     YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
                     AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL
```

Query Match 100.0%; Score 246; DB 1; Length 64957;
Best Local Similarity 100.0%; Pred. No. 1.9e-22;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11513 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 11572

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11573 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCCGACCGCGCGTTCCGC 11632

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11633 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 11692

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11693 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 11752

Qy        241 CACCTC 246
              ||||||
Db      11753 CACCTC 11758
```

```
RESULT 3
E38021
LOCUS       E38021                31422 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Avermectin aglycon synthase gene.
ACCESSION   E38021
VERSION     E38021.1  GI:18626910
KEYWORDS    JP 2000245457-A/2.
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 31422)
  AUTHORS   Omura,S. and Ikeda,H.
  TITLE     Avermectin aglycon synthase gene
  JOURNAL   Patent: JP 2000245457-A 2 12-SEP-2000;
            THE KITASATO INSTITUTE
COMMENT     OS   Streptomyces avermitilis
            PN   JP 2000245457-A/2
            PD   12-SEP-2000
```

```
            THE KITASATO INSTITUTE
COMMENT     OS   Streptomyces avermitilis
            PN   JP 2000245457-A/1
            PD   12-SEP-2000
            PF   24-FEB-1999 JP 1999046961
            PR
            PI   SATOSHI OMURA,HARUO IKEDA
            PC   C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC   (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(11916)
            FT   CDS             (11971). .(30687).
FEATURES              Location/Qualifiers
     source           1. .30690
                      /organism="Streptomyces avermitilis"
                      /db_xref="taxon:33903"
BASE COUNT     5355 a  12455 c   8617 g   4263 t
ORIGIN

  Query Match             100.0%;  Score 246;  DB 6;  Length 30690;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-22;
  Matches  246;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11413 CTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCCTCGCCCTGGTCCGCTCCCAC 11472

Qy         61 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11473 ATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCCCCCCCGACCGCGCGTTCCGC 11532

Qy        121 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11533 GACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCAACCGGCTCTCCCGCACCACC 11592

Qy        181 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11593 GGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCAACCCCACCACCCTCACCCAC 11652

Qy        241 CACCTC 246
              ||||||
Db      11653 CACCTC 11658


RESULT 2
AB032367
LOCUS       AB032367               64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
            Sciences, Kitasato University, Microbial chemistry; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)

FEATURES              Location/Qualifiers
     source           1. .64957
                      /organism="Streptomyces avermitilis"
                      /db_xref="taxon:33903"
     gene             101. .12019
                      /gene="aveA1"
     CDS              101. .12019
                      /gene="aveA1"
                      /note="contains a loading module, Ave module 1 and module
                      2
                      multifunctional polyketide synthase"
                      /codon_start=1
                      /transl_table=11
                      /product="type I polyketide synthase AVES 1"
                      /protein_id="BAA84474.1"
                      /db_xref="GI:5902891"
                      /translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
                      ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
                      AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
                      VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTR
                      MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
                      MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
                      RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
                      DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
                      LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
                      GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
                      MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
                      GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
                      GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
                      RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
                      GDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
                      LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
                      AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAG
                      LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
                      EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
                      VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
                      AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
                      AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
                      ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
                      PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
                      THHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
                      EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
                      WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
                      LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
                      PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
                      LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGAL
                      ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
                      QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
                      TAPVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
                      DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSA
                      ARERERPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
                      FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
                      AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
                      YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
                      GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
                      GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
                      GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
                      NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
                      WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
                      TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
                      SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
                      ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
                      YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
                      VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
                      MTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
                      LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
                      VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
                      NLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
                      TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY
                      VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGG
                      VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
                      RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
                      EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 477.19 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_11413_11658
Perfect score: 246
Sequence: 1 ctggccggccagacacacga.....ccaccctcacccaccacctc 246

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		246	100.0	30690	6	E38020	E38020 Avermectin
2		246	100.0	64957	1	AB032367	AB032367 Streptomy
3		244.4	99.3	31422	6	E38021	E38021 Avermectin
4	c	244.4	99.3	64957	1	AB032367	AB032367 Streptomy
5		152	61.8	12381	6	AX006889	AX006889 Sequence
6		139.4	56.7	30000	6	AX250263	AX250263 Sequence
7	c	138.8	56.4	84985	1	SNA278573	AJ278573 Streptomy
8		134	54.5	30000	6	AX250262	AX250262 Sequence
9		133	54.1	113193	1	AF357202	AF357202 Streptomy
10		133	54.1	123580	1	AF263912	AF263912 Streptomy
11		133	54.1	125401	6	AX211739	AX211739 Sequence
12		130.8	53.2	78210	1	AB070949	AB070949 Streptomy
13		129.8	52.8	65140	6	AX211705	AX211705 Sequence
14		128.2	52.1	43280	1	SFU78289	U78289 Streptomyce
15	c	123.6	50.2	15970	1	AE007061	AE007061 Mycobacte
16	c	123.6	50.2	53450	1	MTV018	AL021899 Mycobacte
17		123.4	50.2	2700	6	AX211733	AX211733 Sequence
18		123.4	50.2	2706	1	AF420594	AF420594 Streptomy
19		123.4	50.2	104326	1	AB070940	AB070940 Streptomy
20		121.8	49.5	1681	6	AR173228	AR173228 Sequence
21	c	121.2	49.3	39314	1	SGR300302	AJ300302 Streptomy
22	c	120.4	48.9	82746	1	AF453501	AF453501 Actinosyn
23		120.2	48.9	25150	1	AB070945	AB070945 Streptomy
24	c	118.8	48.3	24225	1	SC2C4	AL512902 Streptomy
25		118.6	48.2	37948	1	AF079138	AF079138 Streptomy
26		118.6	48.2	41097	1	AF016585	AF016585 Streptomy
27		118.2	48.0	11480	1	STMPKS3ORF	L09654 Streptomyce
28		118.2	48.0	24568	1	AF220951	AF220951 Streptomy
29		118.2	48.0	50937	6	AR159871	AR159871 Sequence
30		118	48.0	27541	6	AX211706	AX211706 Sequence
31		117	47.6	20394	1	SNA132222	AJ132222 Streptomy
32		117	47.6	20394	6	AX067996	AX067996 Sequence
33		116	47.2	32870	1	AF007101	AF007101 Streptomy
34		113.8	46.3	1784	1	AF411573S3	AF411575 Actinomad
35		112.2	45.6	53784	1	AMM223012	AJ223012 Amycolato
36		112.2	45.6	53789	6	A69720	A69720 Sequence 3
37		112.2	45.6	90445	1	AF040570	AF040570 Amycolato
38		111.8	45.4	9769	1	SEERYA	X56107 S. erythrae
39		111.8	45.4	20444	1	SEERYABS	X62569 S.erythraea
40		110.6	45.0	41782	1	AY118081	AY118081 Streptomy
41	c	110.6	45.0	104326	1	AB070940	AB070940 Streptomy
42		109	44.3	17101	1	AB017641	AB017641 Micromono
43		108.8	44.2	20235	1	SERERYAB	M63677 S.erythraea
44		108.8	44.2	20235	6	AR049368	AR049368 Sequence
45		108.8	44.2	20235	6	AR095529	AR095529 Sequence

ALIGNMENTS

RESULT 1
E38020
LOCUS E38020 30690 bp DNA linear PAT 31-JAN-2002
DEFINITION Avermectin aglycon synthase gene.
ACCESSION E38020
VERSION E38020.1 GI:18626909
KEYWORDS JP 2000245457-A/1.
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 30690)
AUTHORS Omura,S. and Ikeda,H.
TITLE Avermectin aglycon synthase gene
JOURNAL Patent: JP 2000245457-A 1 12-SEP-2000;

THIS PAGE BLANK (USPTO)

```
Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              | : :| |:: :  |:|  |:| |:  |   ||| |: :  |  |: |: || ||:|  :
Db        883 CSSSSCCCSSCSCGCSCGCCSCGGSSCCSCGCGCSGSSSGCGCCCSGGSCGCCCGSCGGS 824

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
             :|  ::| | || | :  || |||      : || ::||| |:|    :: | |   :
Db        823 CSCCSSSCCGCGCCGSCSCGCGCCGGCSGGCCSGGCGSSGCGGCSGGCCGSSCCGCSGKS 764

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCG-TCATGGGTGTCGAACTGCGCAACCGCCT 179
             :|| |    |:|: :: :|:|  :| : |::  :  : | | :|  | :|||  :|||:
Db        763 KCGCGCGSGGSCSSSSGSGSCGGSGGSGCSSCSSGCSCGCGSSGSGCGSCGCCCSCGCSC 704

Qy        180 CAGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGC 239
              :|  :: | : | :     |  |||:    |:| |  |  ::  : | :| :::|::|
Db        703 GCSCGCSSGCGSCGCSGSCGCGCCGCCSCGCCCSCCCCSCSSSSCGSSCSSCCSSSCSSC 644

Qy        240 CGCGGTCGCCGCGCGC 255
             |:::| |  | |:|||
Db        643 CSSSGCCCGCCCSCGC 628
```

Search completed: June 18, 2003, 01:00:06
Job time : 333.821 secs

```
Db        253 CGCAGCCCGTGCCAATCATCGCCGA 277


RESULT 14
BE216343
LOCUS       BE216343               901 bp    mRNA    linear   EST 23-OCT-2001
DEFINITION  HV_CEb0010D23f Hordeum vulgare seedling green leaf EST library
            HVcDNA0005 (Blumeria challenged) Hordeum vulgare cDNA clone
            HV_CEb0010D23f, mRNA sequence.
ACCESSION   BE216343
VERSION     BE216343.2  GI:13264851
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 901)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Wei,F., Begum,D.,
            Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi
            ,D.W., Fenton,R.D., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Blumeria infected incompatible (Mla6) seedling
            leaf cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     On Jul 3, 2000 this sequence version replaced gi:8903955.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 393
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 502.
FEATURES             Location/Qualifiers
     source          1. .901
                     /organism="Hordeum vulgare"
                     /cultivar="CI16151 (Mla6)"
                     /db_xref="taxon:4513"
                     /clone="HV_CEb0010D23f"
                     /clone_lib="Hordeum vulgare seedling green leaf EST
                     library HVcDNA0005 (Blumeria challenged)"
                     /tissue_type="seedling green leaf"
                     /lab_host="SOLR"
                     /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: Xho1;
                     C.I. 16151 (Mla6) plants were greenhouse grown in the R
                     Wise lab at Iowa State University, Ames, IA; 7 day old
                     green seedlings were challenged with isolate 5874 (AvrMla6
                     ) of Blumeria graminis f. sp. hordei, and leaves were
                     harvested 20 and 24 hr post-inoculation and snap frozen;
                     uninoculated leaves were harvested 20 hr post-inoculation
                     (Wei, Wise). In the TJ Close lab at the University of
                     California, Riverside, total RNA was prepared from each
                     sample pool, equal quantities of all three RNA pools were
                     combined, poly(A) RNA was purified from the mixture, one
                     primary unamplified cDNA library was made, and 1 million
                     pfu were in vivo excised to give pBluescript SK(-) cDNA
                     phagemids (Choi, Close). Phagemids were plated and picked
                     at the Clemson University Genomics Institute (CUGI) (Begum
                     , Palmer, Frisch, Atkins and Wing). Plasmid DNA
                     preparations, DNA sequencing and sequence analysis were
                     performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates
                     , Rambo, Main). The sequence has been trimmed to remove
                     vector sequence and contains a minimum of 100 bases of
                     phred value 20 or above. For more details on library
                     preparation and sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT      189 a    244 c    313 g    155 t
ORIGIN

  Query Match             16.8%;  Score 43.4;  DB 10;  Length 901;
  Best Local Similarity   50.7%;  Pred. No. 6.9;
  Matches  104;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy         20 CGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGCGGTGCTGAACG 79
              ||  |  |  |  | || |    ||   |||||   ||| |  | ||| ||| |  |   |
Db        132 CGAGGAGGGGGAGGATGATGACGGCAATGGTGAAGGCCGAGGCGATGACGGCGGAGGCGG 191

Qy         80 ACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTTCGCCCACCTCG 139
              |  |   | | || ||| |  |   |||| || |  |          |     | |  || |
Db        192 AGGGGAGCAGCGGGCGGAGGGACGCGGAGGCGGAGCTCAACCTGCCGCCGGGCTTCCGCT 251

Qy        140 GCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGC 199
              |  |  | | | |   |  |||  ||| | ||| |  ||| ||||  |||| | ||
Db        252 TCCACCCCACGGACGACGAGCTCGTCGTGCACTACCTCTGCAGGAAGGTGGCGGGACAGC 311

Qy        200 GGTTGCCCGTGACGCTCATCTTCGA 224
              |  ||||||| |  |||||  |||
Db        312 CGCAGCCCGTGCCAATCATCGCCGA 336


RESULT 15
CNS006XK/c
LOCUS       CNS006XK               935 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066051
VERSION     AL066051.1  GI:4945019
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 935)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .935
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR14N09"
                     /clone_lib="RPCI-98"
                     /note="end : T7"
BASE COUNT      257 a    170 c    162 g     96 t    250 others
ORIGIN

  Query Match             16.7%;  Score 43.2;  DB 17;  Length 935;
  Best Local Similarity   32.4%;  Pred. No. 7.6;
  Matches   83;  Conservative   62;  Mismatches  110;  Indels    1;  Gaps    1;
```

```
  TITLE     Wheat functional genomics - Thatcher Lr1 cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Dr. Sylvie Cloutier
            Cereal Research Centre, Agriculture and Agri-food Canada
            195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
            Tel: (204) 983-2340
            Fax: (204) 983-4604
            Email: scloutier@em.agr.ca
            was cloned directionally, not all sequences generated with reverse
            primer were from the 5' end (same with forward primer and 3' end).
            Average inset size is >2.2 kb
            Plate: 104  row: H  column: 07
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .555
                     /organism="Triticum aestivum"
                     /cultivar="Thatcher Lr1"
                     /db_xref="taxon:4565"
                     /clone="TaLr1104H07"
                     /clone_lib="TaLr1"
                     /tissue_type="Leaf tissue"
                     /dev_stage="14 Days old"
                     /lab_host="E. coli XLOLR"
                     /note="Vector: Lambda ZapII; mass excised in plasmid
                     vector pBK-CMV (Stratagene).; Site_1: EcoRI; Site_2: XhoI;
                     mRNA obtained from wheat NIL Thatcher Lr1 24 hours after
                     inoculation  with leaf rust pathogen Puccinia triticina
                     race BBB carrying the  avirulence gene Avr1."
BASE COUNT       83 a    230 c    162 g     80 t
ORIGIN

  Query Match             16.8%;  Score 43.4;  DB 12;  Length 555;
  Best Local Similarity   48.2%;  Pred. No. 6.6;
  Matches  122;  Conservative    0;  Mismatches  131;  Indels    0;  Gaps    0;

Qy          5 CCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCA 64
              | ||| |    | |||  |||        |||      || | |  ||||    | |   |||
Db          6 CGGCACGAGACGCCAGCCAGAGCATCCCGCCCGCTCTCGCCTCGGTGTCGCCGCCGCCCA 65

Qy         65 TGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGA 124
              |||| ||    ||   | | | |  | ||| | ||||  ||     ||||    |  | |
Db         66 TGGCCGTCGCCAAGAGCAAGGCAGGCAAGCACGCCGCCCCGCTGCTCGGCAAGTACGAGC 125

Qy        125 GTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCA 184
                |  || |||||||  || | || ||      ||| | |       |  ||||  |
Db        126 TCGGCCGCCTCCTCGGCCGCGGCACCTTCGCCAAGGTCTACCACGCGCGCTCCCTCGTCG 185

Qy        185 AGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGG 244
              | |  || | ||    |   |||     ||    ||| |  ||| |  ||||||
Db        186 GCGGCGAGGCCGTGGCCATCAAGGTGCTCGACAAGCCCGAGCTGACCGCCACGGCCGGCA 245

Qy        245 TCGCCGCGCGCCT 257
              | | |||||||  |
Db        246 TGGACGCGCGCGT 258


RESULT 13
BF622916
LOCUS       BF622916                827 bp    mRNA    linear   EST 17-OCT-2001
DEFINITION  HVSMEa0012A21f Hordeum vulgare seedling shoot EST library
            HVcDNA0001 (Cold stress) Hordeum vulgare cDNA clone HVSMEa0012A21f,
            mRNA sequence.
ACCESSION   BF622916
VERSION     BF622916.2  GI:13083794
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 827)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R., Choi,D.W.
            , Fenton,R.D. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Morex cold-stressed seedling shoot cDNA
            library
  JOURNAL   Unpublished (2001)
COMMENT     On Dec 18, 2000 this sequence version replaced gi:11886650.
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 535
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 637.
FEATURES             Location/Qualifiers
     source          1. .827
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEa0012A21f"
                     /clone_lib="Hordeum vulgare seedling shoot EST library
                     HVcDNA0001 (Cold stress)"
                     /tissue_type="Seedling shoot"
                     /lab_host="TJC121"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1;
                     Seeds were surface sterilized then germinated under axenic
                     conditions in the dark at room temperature on filter paper
                     with water, nystatin and cefotaxime in covered
                     crystallization dishes. Five-day old seedlings were
                     incubated at 5oC for 2 days. Shoots were then harvested,
                     total RNA was prepared, poly(A) RNA was purified, one
                     primary unamplified cDNA library was made, and 600000 pfu
                     were in vivo excised to give pBluescript SK(-) cDNA
                     phagemids. These steps were performed in the TJ Close
                     laboratory at the University of California, Riverside
                     (Choi, Close, Fenton). Phagemids were plated and picked at
                     the Clemson University Genomics Institute (CUGI) (Begum,
                     Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations
                     , DNA sequencing and sequence analysis were performed at
                     CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main
                     ). The sequence has been trimmed to remove vector sequence
                     and contains a minimum of 100 bases of phred value 20 or
                     above. For more details on library preparation and
                     sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT      166 a    259 c    292 g    110 t
ORIGIN

  Query Match             16.8%;  Score 43.4;  DB 12;  Length 827;
  Best Local Similarity   50.7%;  Pred. No. 6.8;
  Matches  104;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy         20 CGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGCGGTGCTGAACG 79
              ||  |  |  |  | || |   ||   |||||   ||| |  | ||| ||| |  |   |
Db         73 CGAGGAGGGGGAGGATGATGACGGCAATGGTGAAGGCCGAGGCGATGACGGCGGAGGCGG 132

Qy         80 ACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTTCGCCCACCTCG 139
              | |   | | || ||| | |   |||| || |  |        |      | |  || |
Db        133 AGGGGAGCAGCGGGCGGAGGGACGCGGAGGCGGAGCTCAACCTGCCGCCGGGCTTCCGCT 192

Qy        140 GCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGC 199
              |  |  | | |  |   |||   ||| | ||| |  ||| ||||   |||| | ||
Db        193 TCCACCCCACGGACGACGAGCTCGTCGTGCACTACCTCTGCAGGAAGGTGGCGGGACAGC 252

Qy        200 GGTTGCCCGTGACGCTCATCTTCGA 224
                |  ||||||| |  |||||  |||
```

```
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 1217)
  AUTHORS   Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
            Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
  TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
  JOURNAL   Unpublished (2002)
REFERENCE   2  (bases 1 to 1217)
  AUTHORS   Coe,E.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
FEATURES             Location/Qualifiers
     source          1. .1217
                     /organism="Zea mays"
                     /db_xref="MaizeDB:636162"
                     /db_xref="taxon:4577"
                     /clone="PCO106344"
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT      186 a    452 c    407 g    172 t
ORIGIN

  Query Match           17.1%;  Score 44.2;  DB 11;  Length 1217;
  Best Local Similarity   50.2%;  Pred. No. 4.7;
  Matches  109;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy         39 CCTGGACCTGGTGCGCGCCCACACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCAC 98
              | ||  | | |||  | || | ||   |  |   || | ||||| ||||||
Db        852 CGTGCGCGTCGTGGTCTCCGAGACGGGGTGGCCGTCCAAGGGCGACGCCAACGAGGCGGG 911

Qy         99 CGCGTCGGATGCCGGCCCATCGGCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGG 158
              ||| ||||  |||  | |  |||||     || | |||| | |  |   |    ||   ||
Db        912 CGCTTCGGCGGCCAACGCCGCGGCGTACAACGGCAACCTGGCCCGCCGCGTGCTCTCCGG 971

Qy        159 TGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCAT 218
                ||        |   |||| |  | | |||      |      ||  | | |  |||||
Db        972 CAACGCCGGCACCCCGCGCCGCCCCGACGCCGACATCGACGTGTACCTCTTCGCGCTCTT 1031

Qy        219 CTTCGACCACACCACGCCGGCCGCGGTCGCCGCGCGC 255
              |  |||  ||   | ||| |   ||  | ||| ||||
Db       1032 CAACGAGAACCAGAAGCCCGGGCCGACCTCCGAGCGC 1068


RESULT 11
BE494141
LOCUS       BE494141                 467 bp    mRNA    linear   EST 02-AUG-2000
DEFINITION  WHE1251_F05_K09ZS Secale cereale anther cDNA library Secale cereale
            cDNA clone WHE1251_F05_K09, mRNA sequence.
ACCESSION   BE494141
VERSION     BE494141.1  GI:9660734
KEYWORDS    EST.
SOURCE      rye.
  ORGANISM  Secale cereale
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Secale.
REFERENCE   1  (bases 1 to 467)
  AUTHORS   Anderson,O.D., Butler,E., Chao,S., Choi,D.W., Close,T.J., Fenton
            ,R.D., Gustafson,J.P., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R.,
            Miller,R., Rausch,C.J., Ross,K., Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Anther cDNA library from rye
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .467
                     /organism="Secale cereale"
                     /cultivar="Blanco"
                     /db_xref="taxon:4550"
                     /clone="WHE1251_F05_K09"
                     /clone_lib="Secale cereale anther cDNA library"
                     /tissue_type="Anther"
                     /dev_stage="Adult plant before anthesis"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
                     greenhouse.  Anthers were harvested and pooled  from early
                     meiosis to late meiosis.  The tissue, total RNA, and
                     poly(A)  RNA were prepared (Butler, Ross and Gustafson) at
                     University of  Missouri, Columbia.  A cDNA library was
                     made, and the cDNA clones were  in vivo excised to give
                     pBluescript phagemids in the TJ Close lab  (Choi, Close,
                     Fenton) at the University of California, Riverside.
                     Plasmid DNA preparations and DNA sequencing were performed
                     in the OD  Anderson lab (all other authors)."
BASE COUNT       75 a    167 c    141 g     84 t
ORIGIN

  Query Match           16.9%;  Score 43.6;  DB 10;  Length 467;
  Best Local Similarity   48.8%;  Pred. No. 5.9;
  Matches  118;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy          6 CGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCAT 65
              |||  || || || |   |    |  | |  || |||  | | | |||||||  |||
Db        220 CGCCGTGCCCGCCGCCGGGACCCGTCCCCGCCTACTGCTCATCGCGCGCGCCAGCACGCG 279

Qy         66 GGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAG 125
                | |  |||||| |||||  | || |||  |||||  | | | |||       || |
Db        280 CCGGTTCGTGAACGCCGACGACATCGTGCGGGCCGCGCAGAAGGTCGGGTTCGAGGTGGT 339

Qy        126 TTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAA 185
                .|| |  |   |     |||    |   |   |  || |||  ||||  |  | |
Db        340 GGTCTCGGAGGGCACGCACGAGGTGGCGCCGTTCGCGGAGCTGGCCAACACCTGCGACGC 399

Qy        186 GGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGT 245
              |     |||| |||       ||| |  |  || | | |     |     || ||| || |||
Db        400 GATGCTGGGCGTGCACGGCGCCGGGCTGACCAACATGGTGTTCCTGCCGACGGGCGGGGT 459

Qy        246 CG 247
              |
Db        460 GG 461


RESULT 12
BG909521
LOCUS       BG909521                 555 bp    mRNA    linear   EST 05-JUN-2001
DEFINITION  TaLr1104H07R TaLr1 Triticum aestivum cDNA clone TaLr1104H07 5',
            mRNA sequence.
ACCESSION   BG909521
VERSION     BG909521.1  GI:14317197
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Cloutier,S., Dong,G. and Walsh,A.
```

```
BI553137
LOCUS       BI553137                765 bp    mRNA    linear   EST 05-SEP-2001
DEFINITION  603197725F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5277198 5',
            mRNA sequence.
ACCESSION   BI553137
VERSION     BI553137.1  GI:15440449
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 765)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11700  row: a  column: 07
            High quality sequence stop: 757.
FEATURES             Location/Qualifiers
     source          1. .765
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5277198"
                     /clone_lib="NIH_MGC_95"
                     /tissue_type="hippocampus"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.5 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for  full-length clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health).  Note: this is a NIH_MGC Library."
BASE COUNT      149 a    257 c    257 g    102 t
ORIGIN

  Query Match           17.2%;  Score 44.4;  DB 13;  Length 765;
  Best Local Similarity  50.0%;  Pred. No. 4.1;
  Matches  111;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy         28 CAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGCGGTGCTGAACGACGACGGA 87
              | | ||| |||  || |   || || ||  | ||     |   ||||||  | | |||
Db        393 CCGGCGCAGCTGGTGCAGAAGGCGCACGAGCTCATGTGGTACGTGCTGGTCAAGGACCAG 452

Qy         88 AACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTTCGCCCACCTCGGCTTCGAC 147
              ||  ||   | |   | |  | | | |      |     | ||  |  |  |  |    |
Db        453 AAGAAGATGATCATCTGGTTTCCAGACATGGTGAAAGATGTCATCGGCAGCTACAAGAAG 512

Qy        148 TCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGCGGTTGCCC 207
              |   || | |  ||   | ||||| | |||||| ||  |||  ||   |  || ||| |
Db        513 TGGTGCAGGAGCATCCTCCGGCGCACCAGCCTCATCATCGCCCGGGTGTTCGGGCTGCAC 572

Qy        208 GTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGCC 249
               ||| ||| | |  |   |||||||  |   |   ||| | | |
Db        573 CTGAGGCTAACCAGCCTGCACACCATGGAGTTTGCGCTGGTC 614


RESULT 9
BI600118
LOCUS       BI600118                945 bp    mRNA    linear   EST 07-SEP-2001
DEFINITION  603246515F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:5289003 5',
            mRNA sequence.
ACCESSION   BI600118
VERSION     BI600118.1  GI:15493057
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 945)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
             cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11730  row: m  column: 04
            High quality sequence stop: 847.
FEATURES             Location/Qualifiers
     source          1. .945
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5289003"
                     /clone_lib="NIH_MGC_96"
                     /tissue_type="hypothalamus"
                     /lab_host="DH10B"
                     /note="Organ: brain; Vector: pBluescriptR (modified
                     pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
                     ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',
                     size-selected for  average insert size 2.3 kb and
                     normalized to ROT 5. This is a primary library enriched
                     for full-length  clones and constructed using the
                     Cap-trapper method (Carninci, in preparation).  Library
                     constructed by M. Brownstein (NIMH/NHGRI, National
                     Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      190 a    306 c    306 g    143 t
ORIGIN

  Query Match           17.2%;  Score 44.4;  DB 13;  Length 945;
  Best Local Similarity  50.0%;  Pred. No. 4.2;
  Matches  111;  Conservative    0;  Mismatches  111;  Indels    0;  Gaps    0;

Qy         28 CAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGCGGTGCTGAACGACGACGGA 87
              ||| ||| |||  || |   || || ||  | ||     |   ||||||  | | |||
Db        373 CAGGCGCAGCTGGTGCAGAAGGCGCACGAGCTCATGTGGTACGTGCTGGTCAAGGACCAG 432

Qy         88 AACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTTCGCCCACCTCGGCTTCGAC 147
              ||  ||   | |   | |  | | | |      |     | ||  |  |  |  |    |
Db        433 AAGAAGATGATCATCTGGTTTCCAGACATGGTGAAAGATGTCATCGGCAGCTACAAGAAG 492

Qy        148 TCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGCGGTTGCCC 207
              |   || | |  ||   | ||||| | |||||| |   |||  ||   |  || ||| |
Db        493 TGGTGCAGGAGCATCCTCCGGCGCACCAGCCTCATCCTCGCCCGGGTGTTCGGGCTGCAC 552

Qy        208 GTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGCC 249
               ||| ||| | |  |   |||||||  |   |   ||| | | |
Db        553 CTGAGGCTAACCAGCCTGCACACCATGGAGTTTGCGCTGGTC 594


RESULT 10
AY106411
LOCUS       AY106411                1217 bp    mRNA    linear   HTC 25-MAY-2002
DEFINITION  Zea mays PCO106344 mRNA sequence.
ACCESSION   AY106411
VERSION     AY106411.1  GI:21209489
KEYWORDS    HTC.
SOURCE      Zea mays.
  ORGANISM  Zea mays
```

```
Qy        143 TCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGCGGT 202
              || | |   |  || | ||   |||||||||  |  | ||  ||      |  | |||| |
Db        279 GCGTCGCGCGCGAGGATTTC---CTGCGCAACTACACCGGCTCGGAGCTCGATCCGCGCT 335

Qy        203 TGCCC 207
              || |
Db        336 GGCTC 340


RESULT 6
BG817425
LOCUS       BG817425                 508 bp    mRNA    linear   EST 22-MAY-2001
DEFINITION  EM1_75_G09.b1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BG817425
VERSION     BG817425.1  GI:14188405
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 508)
  AUTHORS   Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
  TITLE     An EST database from Sorghum: developing embryos
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 478
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .508
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Embryo 1 (EM1)"
                     /note="Organ: Embryos germinated for 24 hr; Vector:
                     pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
                     EcoRI; The library was made from poly-A RNA in the cloning
                     vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT       67 a    183 c    174 g     84 t
ORIGIN

  Query Match             17.7%;  Score 45.6;  DB 12;  Length 508;
  Best Local Similarity   52.0%;  Pred. No. 2.2;
  Matches  102;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy          2 TGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACA 61
              || | | || | ||  ||||  || || ||  | |  || |   | |  |  |||| |
Db        244 TGGCGGTGCCGCCCAACACGGCGCCGCTGTTCCAGTACCCGCCGCCGCGGGCCGCGCCGG 303

Qy         62 CCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGG 121
               ||| || ||||  ||  |  ||      |  | || ||||  | | |||||||| |   ||
Db        304 GCATCGCCGTGCGCAAGAAGTACCTGCAGATGGGCGCCGCCGCCGGTGCCGGCGCCGGGG 363

Qy        122 CGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCA 181
              || |  ||| |  | |||||  |      ||||  | |      || ||  | |  ||
Db        364 CGCGCATCGGCGGCGGCGGCTGGGTGGAGTCCATGCGCGCGTCCTCGCCCACGCACGCCA 423

Qy        182 GCAAGGCCACGGGCCT 197
              |   |||| | |  ||
Db        424 GGGCGGCCGCCGCGCT 439
```

```
RESULT 7
AW576621/c
LOCUS       AW576621                 524 bp    mRNA    linear   EST 15-MAR-2000
DEFINITION  UI-HF-BR0p-aka-b-10-0-UI.s1 NIH_MGC_52 Homo sapiens cDNA clone
            IMAGE:3076194 3', mRNA sequence.
ACCESSION   AW576621
VERSION     AW576621.1  GI:7248160
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 524)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. Tissue Procurement: Louis M.
            Staudt, M.D., Ph.D.
             cDNA Library Preparation: M.B. Soares Lab
             cDNA Library Arrayed by: M.B. Soares Lab
             DNA Sequencing by: M.B. Soares Lab
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
             The following repetitive elements were found in this cDNA
            sequence:
             334-381, >GC_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .524
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3076194"
                     /clone_lib="NIH_MGC_52"
                     /tissue_type="lymph"
                     /cell_type="germinal center B cells"
                     /cell_line="MGC85"
                     /lab_host="DH10B (LTI)"
                     /note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
                     Constructed from size fractionated cytoplasmic mRNA
                     (7.4-9.5kb).  Directionally cloned. Cells provided by
                     Louis M. Staudt, Ph.D.  Library preparation by Maria de
                     Fatima Bonaldo, Ph.D. and  M. Bento Soares, Ph.D. "
BASE COUNT       72 a    172 c    194 g     86 t
ORIGIN

  Query Match             17.2%;  Score 44.4;  DB 10;  Length 524;
  Best Local Similarity   50.5%;  Pred. No. 4;
  Matches  108;  Conservative    0;  Mismatches  106;  Indels    0;  Gaps    0;

Qy         32 CGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGCGGTGCTGAACGACGACGGAAACG 91
              ||| || || | | ||| ||  |||||  |  |  |      |  |  |  |||||    ||
Db        380 CGCGGCGCCGGCAGCTGCTGGCCGCCCGCCGCCAGCGCCAGGCGCGCTTCGACGCCGGCG 321

Qy         92 AGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTTCGCCCACCTCGGCTTCGACTCCG 151
              |||  | |  |    |||||  | |    ||||    ||   | |   ||    |
Db        320 AGCTGCCGGACTTCCGTGCCGACACCCGCGCGATCCGCGAAGACGACTGGCGCGTGGCGC 261

Qy        152 TCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGCGGTTGCCCGTGA 211
              | ||   | || ||||  | |||| | |    ||  ||| |||| |     || | ||
Db        260 CCTTGCCCGACGCACTGGTCGACCGGCGCGTGGAGATCACCGGCCCGGTCGAGCGCAAGA 201

Qy        212 CGCTCATCTTCGACCACACCACGCCGGCCGCGGT 245
              | |||||  ||  | ||    |    |||  |||
Db        200 TGATCATCAACGCGCTCAATTCCGGCGCCAAGGT 167


RESULT 8
```

```
  Query Match            18.0%;  Score 46.4;  DB 17;  Length 925;
  Best Local Similarity  13.9%;  Pred. No. 1.5;
  Matches   35;  Conservative  121;  Mismatches   95;  Indels    0;  Gaps    0;

Qy          5 CCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCA 64
              |::|:: : ::: : ::: :: :: | ::: : |::  : |  ::    ||| :|: |:
Db        895 CSSCSSSBSSSSTSSMSSSSBSSSSGSSSSSSGTSSACVKCNASSSCGCCGCGMABCCMC 836

Qy         65 TGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGA 124
               ::::  |: :|: : ::|:|||    | |:| |: ::: ::  ::::::|:    : ::
Db        835 SSSSSCCGSASARGVKVRASGGAGKRGGGSGGASASHSSSSACBSSSSSCSASCWSASSS 776

Qy        125 GTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCA 184
              :   :: :       ||:  ::   ::: :  : | : ::| : :::|  :::|: :|:::
Db        775 SSASSSRSRSGGGAGGSGASSSRSSSSSSSASAGSVVSSASSSSSSCSSSVSCSSVASSM 716

Qy        185 AGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGG 244
              ::: ::: :  ::::  : :|     | ||:: |  :   :::|   :::::::|::::
Db        715 SCSSBSSSSASASSSSSSSSASCASCSCCCTSWSCSCSTSASMSAARSSSSSSSCSSSS 656

Qy        245 TCGCCGCGCGC 255
              : ::: ::::
Db        655 MSASSSASSSS 645


RESULT 4
CNS0052P/c
LOCUS       CNS0052P                844 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR11P16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL056652
VERSION     AL056652.1  GI:4932342
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .844
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR11P16"
                     /clone_lib="RPCI-98"
                     /note="end : TET3"
BASE COUNT      261 a    112 c     92 g     35 t    344 others
ORIGIN

  Query Match            17.8%;  Score 46;  DB 17;  Length 844;
  Best Local Similarity  15.1%;  Pred. No. 1.9;
  Matches   38;  Conservative  122;  Mismatches   91;  Indels    0;  Gaps    0;

Qy          5 CCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCA 64
              :: | |  ::|:|| ::|  : :: ::::| || :: :: |: |: :::| |:| |  :
Db        647 YYCCCCCCYYCSCCBYBCNTSCSYYSYSBCSGCSSSSSSSCSSGSSBYSCCCBCCCBGYC 588

Qy         65 TGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGA 124
                ::: ::::  :: :: :::   :| ||:| ::::::: : :::|  :|: :|  |
Db        587 GCYSBVCSSBSSTBSSVSGSBSSTSSGTGCKCBSSSSSBBCSSBSSCCCYCSSYCTYCYS 528

Qy        125 GTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCA 184
              :  :: |:: |:|:: :  :: :| :: :  |:: |:::    |:::     |: : |
Db        527 SCCBYCCBSGCBTBSYBSCSSGBTGBSSBSGGKSGGKSSGKGGGBKSGGKGYCYCBTYCC 468

Qy        185 AGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGG 244
              :::: |:::::  ::   | :::|:  ::::     :: :  : |:  ::: |::|:::
Db        467 SBSSBCCSSSSSSTSKCCSGGSSSTKCGSSBBGGGSSSKTSGSSSCBSGSSSTGSSGSSK 408

Qy        245 TCGCCGCGCGC 255
              ::::: :::::
Db        407 KYSYSTSSSSS 397


RESULT 5
AZ934718
LOCUS       AZ934718                594 bp    DNA     linear   GSS 24-APR-2001
DEFINITION  BJ__Ba0002F10r B. japonicum BAC library Bradyrhizobium japonicum
            genomic, DNA sequence.
ACCESSION   AZ934718
VERSION     AZ934718.1  GI:13776778
KEYWORDS    GSS.
SOURCE      Bradyrhizobium japonicum.
  ORGANISM  Bradyrhizobium japonicum
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE   1  (bases 1 to 594)
  AUTHORS   Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Goicoechea
            ,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
  TITLE     A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
            genome
  JOURNAL   Genome Res. 11 (8), 1434-1440 (2001)
  MEDLINE   21376150
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Class: BAC ends
            High quality sequence start: 22
            High quality sequence stop: 549.
FEATURES             Location/Qualifiers
     source          1. .594
                     /organism="Bradyrhizobium japonicum"
                     /strain="USDA110"
                     /db_xref="taxon:375"
                     /clone_lib="B. japonicum BAC library"
                     /lab_host="E. coli"
                     /note="Vector: pIndigo536; Site_1: HindIII"
BASE COUNT      130 a    192 c    171 g    100 t      1 others
ORIGIN

  Query Match            17.8%;  Score 45.8;  DB 17;  Length 594;
  Best Local Similarity  56.8%;  Pred. No. 2;
  Matches  105;  Conservative    0;  Mismatches   77;  Indels    3;  Gaps    1;

Qy         23 AGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGCGGTGCTGAACGACG 82
              || || || |||||| ||| | ||| | ||| | ||| | | | | || | | |||
Db        159 AGGAGAAGACGCTGCGCCTCAACCAGTCGCGTATCGACAGTCTCGTCGAGCAGCTCTACG 218

Qy         83 ACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTTCGCCCACCTCGGCT 142
              ||   ||| |||||  ||  ||| ||| ||||     |  | || ||||| ||  |  |
Db        219 ACATCAACAAGCGCCTCGTCTCGCATGAGGGCCGCCTGATGCGTCTCGCCGACAGCCACG 278
```

```
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .925
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone="BACR19D16"
                      /clone_lib="RPCI-98"
                      /note="end : TET3"
BASE COUNT      120 a     61 c     61 g    172 t    511 others
ORIGIN

  Query Match             18.6%;  Score 48;  DB 17;  Length 925;
  Best Local Similarity   16.0%;  Pred. No. 0.7;
  Matches   41;  Conservative  126;  Mismatches   85;  Indels    4;  Gaps    1;

Qy          5 CCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCA 64
              ||:|:::  ::::: ::| :  :: : | : :: | |:  :: :  ::::: :: : ::
Db        593 CCSCSSYCCSSSBSSSKCSSTSBSCSCCCSSKSVCGTSCSSSSSCSSSSSSTSSSSTSSS 652

Qy         65 TGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGA 124
              |:::::|: :  :: :  :::  |:| |: :|  |:|  |:  :::::::  |:::::
Db        653 TSKSSSSGSSSSSSSSYTTSKSTSASGSGSWSAGGGSGSTGSTSSSSSSSSTSTSSSSVS 712

Qy        125 GTTTCGCCCACCTCGGCTTCGACTCCGTCAT----GGGTGTCGAACTGCGCAACCGCCTC 180
              :  ::: ::  :::::   :: :|::::: |    ::: ::::  |  | |  ::::: :
Db        713 SGSKSSTBSSGSBSSSGSSSSSSTSSBBSCTSTSSSSSSSYSSSTCSCCTCCCSYSYSSS 772

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              ::  : :  : :|:: :::|| ::: :|: | : | :: ||   : : |: : :|||::
Db        773 TSSSSSTSWGSTSGSSSSSVGTSSSSDSTSTCCSCCCYMCTCCSTYBMBCYTSTSCGGSS 832

Qy        241 GCGGTCGCCGCGCGCC 256
              :::|: |: :|||| |
Db        833 SSSGKGGVTKCGCGGC 848


RESULT 2
AZ934566
LOCUS       AZ934566                 591 bp    DNA     linear   GSS 24-APR-2001
DEFINITION  BJ__Ba0002F19f B. japonicum BAC library Bradyrhizobium japonicum
            genomic, DNA sequence.
ACCESSION   AZ934566
VERSION     AZ934566.1  GI:13776626
KEYWORDS    GSS.
SOURCE      Bradyrhizobium japonicum.
  ORGANISM  Bradyrhizobium japonicum
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Goicoechea
            ,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
  TITLE     A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
            genome
  JOURNAL   Genome Res. 11 (8), 1434-1440 (2001)
  MEDLINE   21376150
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Class: BAC ends
            High quality sequence stop: 539.
FEATURES              Location/Qualifiers
     source           1. .591
                      /organism="Bradyrhizobium japonicum"
                      /strain="USDA110"
                      /db_xref="taxon:375"
                      /clone_lib="B. japonicum BAC library"
                      /lab_host="E. coli"
                      /note="Vector: pIndigo536; Site_1: HindIII"
BASE COUNT      130 a    196 c    176 g     89 t
ORIGIN

  Query Match             18.4%;  Score 47.4;  DB 17;  Length 591;
  Best Local Similarity   57.3%;  Pred. No. 0.9;
  Matches  106;  Conservative    0;  Mismatches   76;  Indels    3;  Gaps    1;

Qy         23 AGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGCGGTGCTGAACGACG 82
              ||  | |||||||||  |||  ||| || |||   | |||   | |  | || |  | |||
Db        107 AGGTGAAGTCGCTGCGCCTCAACCAGGCGCGTATCGACAGTCTCGTCGAGCAGCTCTACG 166

Qy         83 ACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTTCGCCCACCTCGGCT 142
              ||   ||| |||||  ||  ||| |||  ||||    |  | || ||||| ||  |  |
Db        167 ACATCAACAAGCGCCTCGTCTCGCATGAGGGCCGCCTGATGCGTCTCGCCGACAGCCACG 226

Qy        143 TCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGCGGT 202
              || | |   |  || | ||   |||||||||  |  | ||  ||     |  | ||| |
Db        227 GCGTCGCGCGCGAGGATTTC---CTGCGCAACTACACCGGCTCGGAGCTCGATCCGCGCT 283

Qy        203 TGCCC 207
              ||| |
Db        284 GGCTC 288


RESULT 3
CNS0091P/c
LOCUS       CNS0091P                 925 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .925
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone="BACR19D16"
                      /clone_lib="RPCI-98"
                      /note="end : TET3"
BASE COUNT      120 a     61 c     61 g    172 t    511 others
ORIGIN
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 330.73 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_1096_1353
Perfect score: 258
Sequence: 1 ctgaccgcgctgtcccccac.....ccgcggtcgccgcgcgcctt 258

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	48	18.6	925	17	CNS0091P	AL053013 Drosophil
	2	47.4	18.4	591	17	AZ934566	AZ934566 BJ__Ba000
c	3	46.4	18.0	925	17	CNS0091P	AL053013 Drosophil
c	4	46	17.8	844	17	CNS0052P	AL056652 Drosophil
	5	45.8	17.8	594	17	AZ934718	AZ934718 BJ__Ba000
	6	45.6	17.7	508	12	BG817425	BG817425 EM1_75_G0
c	7	44.4	17.2	524	10	AW576621	AW576621 UI-HF-BR0
	8	44.4	17.2	765	13	BI553137	BI553137 603197725
	9	44.4	17.2	945	13	BI600118	BI600118 603246515
	10	44.2	17.1	1217	11	AY106411	AY106411 Zea mays
	11	43.6	16.9	467	10	BE494141	BE494141 WHE1251_F
	12	43.4	16.8	555	12	BG909521	BG909521 TaLr1104H
	13	43.4	16.8	827	12	BF622916	BF622916 HVSMEa001
	14	43.4	16.8	901	10	BE216343	BE216343 HV_CEb001
c	15	43.2	16.7	935	17	CNS006XK	AL066051 Drosophil
	16	43	16.7	691	10	BE214360	BE214360 HV_CEb000
	17	42.8	16.6	637	14	BQ640195	BQ640195 he25d09.y
	18	42.8	16.6	651	13	BG913908	BG913908 602812419
	19	42.8	16.6	678	10	BE408592	BE408592 601301983
c	20	42.8	16.6	728	17	AZ933916	AZ933916 BJ__Ba000
	21	42.8	16.6	734	12	BG707799	BG707799 602671142
	22	42.8	16.6	740	12	BG708101	BG708101 602671552
	23	42.8	16.6	748	10	BE274823	BE274823 601122771
	24	42.8	16.6	872	9	AU122140	AU122140 AU122140
	25	42.8	16.6	880	9	AU119522	AU119522 AU119522
	26	42.8	16.6	883	12	BG761724	BG761724 602717918
	27	42.8	16.6	903	9	AL543324	AL543324 AL543324
	28	42.8	16.6	941	14	BQ879183	BQ879183 AGENCOURT
	29	42.8	16.6	944	13	BI757009	BI757009 603028730
	30	42.8	16.6	1151	14	BM924919	BM924919 AGENCOURT
	31	42.6	16.5	675	9	AU126183	AU126183 AU126183
	32	42.4	16.4	566	14	BQ752993	BQ752993 WHE4121_F
	33	42.4	16.4	567	9	AL504599	AL504599 AL504599
c	34	42.2	16.4	557	14	BQ762235	BQ762235 EBro01_SQ
c	35	42.2	16.4	681	13	BM347502	BM347502 MEST278-D
	36	42.2	16.4	956	12	BG837420	BG837420 Zm10_07b1
	37	42	16.3	926	12	BG321314	BG321314 Zm04_04d1
	38	41.8	16.2	477	12	BG354194	BG354194 947032E01
	39	41.8	16.2	857	17	AG061653	AG061653 Pan trogl
	40	41.8	16.2	1101	17	CNS017SY	AL108460 Drosophil
	41	41.6	16.1	390	10	BE588154	BE588154 WHE0657_C
	42	41.6	16.1	625	13	BM440334	BM440334 pgr1n.pk0
	43	41.6	16.1	629	13	BM487133	BM487133 pgm2n.pk0
	44	41.6	16.1	634	13	BM487449	BM487449 pgm2n.pk0
	45	41.6	16.1	676	9	AJ444452	AJ444452 AJ444452

ALIGNMENTS

RESULT 1
CNS0091P
LOCUS CNS0091P 925 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL053013
VERSION AL053013.1 GI:4934461
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 925)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial

```
Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
Db        769 NGNNNGNNGNNGGGGNNGGGGGGGGGGGGGGGGGGGGGNGNNGGGGGGGNGGGGGGGGGGG 828

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
Db        829 GGGGGGGGGGGGGGGGGGGGGGGNNGGGGGGGGGGGGGGGGGGGGGGGGGGG 882


RESULT 15
BI416517
LOCUS       BI416517                 897 bp    mRNA    linear   EST 15-AUG-2001
DEFINITION  hasp001xd10f Heterobasidion annosum - Scots pine infection stage
            subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
            annosum cDNA clone hasp001xd10f, mRNA sequence.
ACCESSION   BI416517
VERSION     BI416517.1  GI:15187540
KEYWORDS    EST.
SOURCE      Pinus sylvestris/Heterobasidion annosum.
  ORGANISM  Pinus sylvestris/Heterobasidion annosum
            Eukaryota; mixed EST libraries.
REFERENCE   1  (bases 1 to 897)
  AUTHORS   Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
  TITLE     Expressed sequence tags of randomly selected cDNA clones from the
            interaction of the root rot fungus (Heterobasidion annosum) with
            seedling roots of Scots pine (Pinus sylvestris)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Fred O. Asiegbu
            Dept. of Forest Mycology & Pathology
            Swedish University of Agriculture, Box 7026,S-750 07, Uppsala,
            Sweden
            Tel: +46 18 67 15 98
            Fax: +46 18 30 92 45
            Email: Fred.Asiegbu@mykopat.slu.se
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .897
                     /organism="Pinus sylvestris/Heterobasidion annosum"
                     /db_xref="taxon:169015"
                     /clone="hasp001xd10f"
                     /clone_lib="Heterobasidion annosum - Scots pine infection
                     stage subtraction cDNA library (hasp)"
                     /dev_stage="Seedling roots of scots pine were infected for
                     6 days with H. annosum"
                     /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive
                     hybridization cDNA library was constructed from  scots
                     pine roots infected for 6-days with mycelia of
                     Heterobasidon  annosum (FP5)."
BASE COUNT       12 a     21 c    709 g     47 t    108 others
ORIGIN

  Query Match           17.5%;  Score 93.4;  DB 13;  Length 897;
  Best Local Similarity   45.7%;  Pred. No. 1.9e-08;
  Matches  244;  Conservative    0;  Mismatches  290;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
Db        150 GGGGGGGGGNGGGGGNNNNNNNGNNGGGGGGGGGGGGGGGGGGGGNGGGGGNGNNGNNGG 209

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
Db        210 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGTGGGGGGGGGGTNGGGGGGG 269

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
Db        270 GGGGGGGGGNNNNNGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGTGGGGGTTGTGGNG 329

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
Db        330 GTNGGGGNGGGGGGGGGGNGGGGGGGTGTGGGGGGTGGGGGGGGGGGNNGGGGGGGGGG 389

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
```

```
Db        390 GGGGGGGGGGGTGTGGGGGGGGGGGGGGGGGGGGGGGGGGGTGGGGGGGGGGNGTGGG 449

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
Db        450 GGGGGGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGNGGTGNGGNGGGGNNGGGGGGNGGG 509

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
Db        510 GGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGTGNGGGGGGGGGGGGG 569

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
Db        570 GTGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGTGGGGGGGGGGGGGGGGGGGGGG 629

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
Db        630 TGNGGTGGGGGGGGGGGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 683
```

Search completed: June 18, 2003, 01:00:31
Job time : 686.626 secs

```
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLCM1978  row: d  column: 07
           High quality sequence stop: 289.
FEATURES             Location/Qualifiers
     source          1. .1184
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5470806"
                     /clone_lib="NIH_MGC_41"
                     /tissue_type="amelanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT       66 a     92 c    843 g     42 t    141 others
ORIGIN

  Query Match             17.6%;  Score 93.8;  DB 14;  Length 1184;
  Best Local Similarity   43.1%;  Pred. No. 1.7e-08;
  Matches  230;  Conservative    0;  Mismatches  304;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              |||  || |  || |  ||| |||  ||| | |  ||| | || || || | ||     |
Db        567 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGG 626

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |   | | | | |||| | |||    |  |  || | |     | ||   ||  |
Db        627 GGGNGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNNNGGGGGGGGGGGGGGGN 686

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || | || |||   | ||| || |  | ||  | |  |  || |  || | ||  .  | | |
Db        687 GGGGNGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNGGNGNGGGGGNGGGGGGNGGGG 746

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              | |  | ||| | | || ||| || |     ||  |     ||| |     |  ||| |
Db        747 GGGGGGGGGGGNGGGGGGGGGGNGGGGNNGGGGGGGNNNNGGGGGGNGGGGGGGGGGGGG 806

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |  |  ||| |  | |  | ||  || | |  ||  |  | |  |     |
Db        807 GGGGGGGGGGGGGGNGGGNNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGN 866

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |      | ||  |||   || |   | ||    || |   |           |  ||
Db        867 NNGGGNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNNGGNNNNGGGGGGGGGGGG 926

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              | |  |  |  |||       |  |  |  |  | |  |         ||  | ||||
Db        927 GGGGGGNGGNNGGGGGNNNNGGGGNGGNGGGNNGGGGGGGGNNNNGGGGGGGGGGGGGGG 986

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              ||||  ||   ||   ||      |  | ||     |          ||  | |  || |
Db        987 GGGGGGNGGGGNNGGGGNGGGGGNGNNGGGGNGGGGGGGGGGNNNNGGGGGGGGNGGGGG 1046

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              || || ||| ||  ||   |  ||||| |  |  || | |   |  | ||
Db       1047 NGGGGGGGGGGGGGGGGNGGGNGGGGGGGGGGGGGGGNGGGGGGNNNGGGGGGG 1100


RESULT 14
AZ189327
LOCUS       AZ189327                 888 bp    DNA     linear   GSS 30-AUG-2000
DEFINITION  SP_1014_A1_F05_T7A Strongylocentrotus purpuratus, purple sea urchin
            , sperm genomic BAC library Strongylocentrotus purpuratus genomic
            clone Plate=1014 Col=9 Row=K, DNA sequence.
ACCESSION   AZ189327
VERSION     AZ189327.1  GI:8372605
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 888)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and
            Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 1014  row: K  column: 9
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 888.
FEATURES             Location/Qualifiers
     source          1. .888
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="Plate=1014 Col=9 Row=K"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
BASE COUNT       52 a     37 c    606 g     83 t    110 others
ORIGIN

  Query Match             17.5%;  Score 93.4;  DB 17;  Length 888;
  Best Local Similarity   44.6%;  Pred. No. 1.9e-08;
  Matches  238;  Conservative    0;  Mismatches  296;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              |||  || |  || |  ||| |||  ||| | |  ||| | || || || | ||     |
Db        349 GGGNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 408

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  || | | | |||| | |||    |  | || | |  || | ||   ||  |
Db        409 GGGGGGGGGGGGGGGGGGGGGGNNNGGNGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGN 468

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              | | | |||   | ||| || |  || | |  ||||| || || | ||   |  | |
Db        469 NGNGNNGNGGGGGGGGGGGGGGGGGGGGNGGNGGGGGGGGGGGGGGGGGGGGGGGGGG 528

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              | |  | ||||  | ||  || || ||   ||  |  || ||| |     |  || |
Db        529 GGGGGGGGGGGGNNGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGNGNNNGNGGGNGGG 588

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |  |  ||| |  | |  | ||  || | |  ||  |  | |  |     |
Db        589 GGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGGNGGGGGGGNGGGGGGGGGGGGGGG 648

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |   | | ||  |||   |  |   | ||    || ||  |  |         |  ||
Db        649 GGGGNGGGGNGGGGGGGGGGGNGGGGGGGGGNGGNGGGGGGGGGGGGGGGGGGGGGGG 708

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              | |  |  |  ||||  |  |  |  || |  |  |         ||  | ||
Db        709 GGGGNGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGNNGGGGGNGGGNGGGG 768
```

```
ORIGIN

  Query Match             17.7%;  Score 94.6;  DB 14;  Length 993;
  Best Local Similarity   44.6%;  Pred. No. 1.2e-08;
  Matches  238;  Conservative    0;  Mismatches  296;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||| || | |  | ||| ||| ||| | |  ||| | || || || |  |      |
Db        385 GGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGG 444

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  || | | |||| |  ||    |  |  ||    |  || | ||   ||  |    |
Db        445 GGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGNGGGNGGGGGGGGGGGGGGGGGG 504

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || | || |||   | ||| || |  ||  | |  ||||  |  || | ||   | | |
Db        505 GGNGGGGGGGGGNGGGGGGGNGGGGGNGGGGGGGGNGGGGGGGNNGGNGGGGGGGNGGGG 564

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              | | | | ||| | || ||| || ||   ||  |  || ||| |      | ||| |
Db        565 GGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGGNNNGGGGGGGGG 624

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |  |  | |    | |  | || || | |  ||  | | | |  |
Db        625 GGGGGGGGGNGNNGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGNGGG 684

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |  | | ||  ||| |  |  | ||  || ||      |      | ||
Db        685 GNGNNNGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGNGGGGGGGGNGGGGGGG 744

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |  |  | ||||  || |  |  | ||  |  |  |  |  || | ||||
Db        745 GGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGNNGNGGNGNNGNGGGGGNGGGGGGNN 804

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              |||  ||  |||  |||        |  | |        |  |    ||    | || |
Db        805 GNNGGGGGGGGGGGGGGGGGGGGGGGGNGGGGNGGGGGGGGGGGGGGGGNGGGGGGG 864

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              || ||  ||  | |||   | |||||  |  || | ||||   | ||
Db        865 GGGGGGGGGNNGGGNGGGGGGNNGGGGGGGGGGGGGGNGGGGGGGGGNGGGGGG 918


RESULT 12
BM468944/c
LOCUS       BM468944               1089 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6447816 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:5587245
            5', mRNA sequence.
ACCESSION   BM468944
VERSION     BM468944.1  GI:18517986
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1089)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12356  row: g  column: 22
            High quality sequence stop: 329.
FEATURES             Location/Qualifiers
     source          1. .1089
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5587245"
                     /clone_lib="NIH_MGC_92"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      131 a    692 c     87 g     79 t    100 others
ORIGIN

  Query Match             17.6%;  Score 93.8;  DB 13;  Length 1089;
  Best Local Similarity   45.3%;  Pred. No. 1.6e-08;
  Matches  242;  Conservative    0;  Mismatches  292;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||| || |   |  |||||| ||| | |  |||   || || || | ||      |
Db        905 GGGGGGGGGGGNNNGGGGGGTGGGGGGGGGGGGGGGGNGGGNGGGGGGGGGGGGNG 846

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  || |   | || | |||         |  || | |  || | ||    |  |    |
Db        845 GGNGGNGGNNGNGNGGGGGGGGGGGGTTGGGGGGGGGGGGGGGGGGGGNNNGGNGGNNGG 786

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || | || |||   | ||| || |  ||  | |  ||||| || || | ||   | | |
Db        785 GGGGGGGGGGNGNNGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGG 726

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              | | | |||| | || ||  || ||   ||  | || ||| |      |  ||| |
Db        725 GGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGG 666

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |  |  ||| | |  | || | |  | |  || | | |  |  |  |  |  |
Db        665 GGGGGGGGGGGGGNGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGNGNGGNGGGGGG 606

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |  | | ||  ||| |  |  | ||  || ||      |      | ||
Db        605 GGGGGGGGGGGGGGGGGNGNGGGNGGGGGGGGGNGGGGGGGGGGGNGGNGGGGGGGGGG 546

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |  |  | ||||  || |  |  | ||  |  |  |  |  || | ||||
Db        545 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGNNGGGGGGGGGGGGGGGGGG 486

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              ||||  ||  |||  |||        |  | |        |  |    ||    | || |
Db        485 GGGGGGGGGGGNGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 426

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              || ||  ||  | |||   | |||||  |  || | ||||   | ||
Db        425 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNGGNGGGGGGGGGGGGGNNG 372


RESULT 13
BM911719
LOCUS       BM911719               1184 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6640314 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5470806
            5', mRNA sequence.
ACCESSION   BM911719
VERSION     BM911719.1  GI:19362098
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1184)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
```

```
Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |     | ||  |||   || |   | ||    ||  || ||         |  ||
Db        314 GGGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGCGAGGCGCGGGGGGGGGGGGGG 373

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              | |   |    ||||   || || |  || |   | |             ||  ||||||
Db        374 GGGGGGGGCGGGGGGGGGGGGCGGGGGGGGGGGGCGGGCGGGGGGGGGGGGGCGGGGGGG 433

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              ||||  ||| |||  |||       | |  |      |  |  :  | ||  |  || |
Db        434 GGGGGGGGGGCGGGGGGGGGGCGGGGGGCCGGAGGGGGGGGGGSGCGGCGGGGGGGGG 493

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                ||| ||| || || |||   |  ||| | |  || || | |||||  | ||
Db        494 CCGCGGGGGCCGGGGGGGGGGGGGGGGGCGGGGGGCGGGGGGGGGGGGGGGG 547


RESULT 10
BQ941425
LOCUS       BQ941425                937 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_8820986 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6422487
            5', mRNA sequence.
ACCESSION   BQ941425
VERSION     BQ941425.1  GI:22356903
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 937)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2599  row: i  column: 16
            High quality sequence stop: 290.
FEATURES             Location/Qualifiers
     source          1. .937
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6422487"
                     /clone_lib="NIH_MGC_18"
                     /tissue_type="large cell carcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT       61 a     91 c    614 g     53 t    118 others
ORIGIN

  Query Match             17.8%;  Score 94.8;  DB 14;  Length 937;
  Best Local Similarity   43.9%;  Pred. No. 1.1e-08;
  Matches  234;  Conservative    0;  Mismatches  299;  Indels    0;  Gaps    0;

Qy          2 GGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTGG 61
              ||  || |   | |  ||| |||  ||| | |  ||| | || ||  | | ||     ||
Db        401 GGGNGGGGGGNGGGNNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNGGGGGGGGGGGG 460

Qy         62 CTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCGG 121
                || | | |||| | ||              | |  || | ||   ||  |    ||
```

```
Db        461 GGGGGGGGGGGGGGGGGGGGNGGNNNNNNGNNNGGGGGGGGGGGGGGGGGGGGGGGGG 520

Qy        122 GTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCGT 181
              | | || |||   | ||| || |  ||  | |  ||| | || || | ||   |  |
Db        521 GGGGGGGGGGNGGGGGGGNGGGGGGGGGGGGGNNGGNGGGGGGGGGGGGGGGGGNNN 580

Qy        182 GTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGTC 241
              | |  | || || | || ||| || ||   ||  |  || ||  |      |  ||| |
Db        581 GGGNNGGGGNGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGNGGGGGGGNNGGGGGNG 640

Qy        242 CGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTCA 301
                  ||| |  | |      ||  || | |  ||  |  | |  |   |
Db        641 NNNGNGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGNG 700

Qy        302 CGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGATG 361
              |   | | ||  |||   || |  | ||    || ||  |  |      |  ||  |
Db        701 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 760

Qy        362 AGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATCC 421
              |  | | ||||   || |  |  || |  | |  |          ||  | ||||
Db        761 GGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGNNGGGGNNGGNGGGGGGGGGG 820

Qy        422 TGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCGT 481
              ||||  ||  |||  |||         |   ||      |  |    ||  | |  || |
Db        821 GGGGGGNGGGGGGGGGGGGNNNGGGGGGNNGGGGGGGGGGGGGGGGGGGNGGGGNGGGG 880

Qy        482 ACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                |  || ||| || |||   |   |||| |  |   | |  |||   | ||
Db        881 GNNGNGGGGGGGGGGGGGGGNGGNNNGGGGGGGGGGGNGGGNNGGGNNGGGGG 933


RESULT 11
BQ958704
LOCUS       BQ958704                993 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_8932917 NCI_CGAP_Mam2 Mus musculus cDNA clone
            IMAGE:6489144 5', mRNA sequence.
ACCESSION   BQ958704
VERSION     BQ958704.1  GI:22374182
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 993)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14036  row: o  column: 01
            High quality sequence stop: 271.
FEATURES             Location/Qualifiers
     source          1. .993
                     /organism="Mus musculus"
                     /strain="FVB/N-3"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6489144"
                     /clone_lib="NCI_CGAP_Mam2"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT       82 a     31 c    683 g     74 t    123 others
```

```
ACCESSION   BQ939440
VERSION     BQ939440.1  GI:22354918
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1089)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2670  row: g  column: 02
            High quality sequence stop: 327.
FEATURES             Location/Qualifiers
     source          1. .1089
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6484345"
                     /clone_lib="NIH_MGC_40"
                     /tissue_type="carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
                     Site_2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed  by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California,  Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life  Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT       73 a     59 c    712 g     81 t    164 others
ORIGIN

  Query Match             17.9%;  Score 95.4;  DB 14;  Length 1089;
  Best Local Similarity   44.9%;  Pred. No. 8.2e-09;
  Matches  240;  Conservative    0;  Mismatches  294;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||| || |  || |  ||| ||| ||| | |  ||| | || || || | ||      |
Db        512 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 571

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              | || | | |||| | |||    | |  || | |  || | ||   ||  |    |
Db        572 GNGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 631

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || | || |||   | ||| || |  ||  | |  ||||| || || | ||   | | |
Db        632 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 691

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              |         | | | || ||| || |     |   | || ||| |     | |||  |
Db        692 GGGNNNNGNGNNGGGGGGGGGGGGGGGNNNNGNNNGGGGGGGGGGGGGGGGGGGGNGG 751

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |  | |||  | | |  | ||  || | |  ||        | |  |
Db        752 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNNNGNNGGNGGNGGNNNGNGN 811

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |  | | ||  |||   ||     | ||    || ||  |           |  ||
Db        812 NGGGGGGGGGGGGGGGGNNGGGGNGGGGGGGNGGGGGGGGGGGGGNNGGGGNNGNNGGGG 871

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              | | | | | ||  || | |     || | | | |          || | ||||
Db        872 GGGGGGNGGGGNGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 931

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              ||||  ||  |||  |||          | | |     | |     ||    | || |
Db        932 GGGGGGGGGGGGGGGGGGGGNGGGGNGGGGGGNGGGGGGGGGGGGGGGGGGNGGGGGNG 991

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              || || ||| || |||   | ||||| |  |  || | |||||  | ||
Db        992 GGGNGGGGGGGGGGGGGGGGGGGGNGGGGGGGGNGGGGGGGGGGGGGGGNGG 1045


RESULT 9
AL572700
LOCUS       AL572700                855 bp    mRNA    linear   EST 16-FEB-2001
DEFINITION  AL572700 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DI033YC01 3
            prime, mRNA sequence.
ACCESSION   AL572700
VERSION     AL572700.1  GI:12931222
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 855)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .855
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="CS0DI033YC01"
                     /clone_lib="LTI_NFL006_PL2"
                     /tissue_type="placenta"
                     /note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer.  Five prime end
                     enriched, double-stranded cDNA was digested with Not I and
                     cloned into the Not I and Eco RV sites of the pCMVSPORT 6
                     vector. Library was normalized. Library was constructed by
                     Life Technologies. Contact : Feng Liang Life Technologies,
                     a division of Invitrogen 9800 Medical Center Drive
                     Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
                     Email : fliang@lifetech.com URL :
                     http://fulllength.invitrogen.com"
BASE COUNT       67 a     95 c    629 g     43 t     21 others
ORIGIN

  Query Match             17.8%;  Score 95.2;  DB 9;  Length 855;
  Best Local Similarity   47.4%;  Pred. No. 8.8e-09;
  Matches  253;  Conservative    9;  Mismatches  272;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              |||  || || ||  || ||| |||  ||| | |  ||| | || || || | ||      |
Db         14 GGGGGGGGGTGGGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 73

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              | || | | |||||| | |:|    | |  || | |  | | || | | :|  |    |
Db         74 GGGGGGGGGGGTGGGGGGGGSGGGGGGGGGGGGGGGGGGGGAGGGGGGGCGSGGGGGGAGG 133

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              | | || |||   ||| | || |  ||  | |  ||:|| || || | ||   | ||||
Db        134 CGGGGGGCGGGGGGGCGTGAGGGGGGGGGGGGGGGGGSGGGGGTGGGGGGGGGCGGTGCG 193

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              | |  | |||||   :| ||| |  ||   ||  | :|| ||| |   | ::| || |
Db        194 GGGGGGGGGGGGGTTGSGGGGGGGCGGGGGGGGGGGGKGGGGGGGGGGGCGSSCCGGCGG 253

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |  | |  || | |  || | || ||| ||  | | | | |  |  |:
Db        254 GGGGGGGGCGGGGGGGCGGGGGCTGGCGGGGTGGGGGGGGGGGGGGGCGGGSGGGGGGG 313
```

/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 26 a 54 c 746 g 30 t 233 others
ORIGIN

Query Match 17.9%; Score 95.8; DB 14; Length 1089;
Best Local Similarity 46.8%; Pred. No. 6.9e-09;
Matches 250; Conservative 0; Mismatches 284; Indels 0; Gaps 0;

```
Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              |||  ||  |  ||  |  ||| |||  ||| | |  ||| | ||  ||  ||  | ||     |
Db        374 GGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 433

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  ||  |  |  |||| |  |||     |  |  ||  |  |  ||  |  ||   ||  |  |     |
Db        434 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 493

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              ||:|  ||  |||    |  |||  ||  |  ||  |  |||||  ||  ||  |  ||    |  |  |
Db        494 GGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 553

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              |  |  |||||  |  ||  |||  ||  ||    ||  |  ||  |||  |       |  |||  |
Db        554 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 613

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |  |  |||  |  |  |  ||  ||  |  |  |  |  |  |  |
Db        614 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGNNNNGG 673

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |  |  |  ||  |||  ||  |  |  ||  ||  |  |  |  |  ||
Db        674 NGGGGGGGNGGGGGGGGGGGGGGGGGGNGNGGGGNGGGGGGGGGGNGGGGGGGGGGGGG 733

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |  |  |  |  ||||  ||  |  |  ||  |  |  |  ||  |  ||||
Db        734 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGNGG 793

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              ||||  ||  ||  |||  |  |  ||  |  |  ||  |  |  ||
Db        794 GGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGNGGGGGGNNGGGGGNGGGGGGGNG 853

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||  ||  |||  ||  |||   |  |||||  |  |  ||  |  |||||  |  ||
Db        854 GGNNGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGNGGNGNGGGGGGGGGG 907
```

RESULT 7
BQ943816
LOCUS BQ943816 1136 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT_8773466 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:6372516 5', mRNA sequence.
ACCESSION BQ943816
VERSION BQ943816.1 GI:22359294
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1136)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP/Gazdar
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2548 row: g column: 13
High quality sequence start: 54
High quality sequence stop: 323.

FEATURES Location/Qualifiers
source 1. .1136
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6372516"
/clone_lib="NIH_MGC_18"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 80 a 268 c 669 g 65 t 54 others
ORIGIN

Query Match 17.9%; Score 95.6; DB 14; Length 1136;
Best Local Similarity 48.4%; Pred. No. 7.5e-09;
Matches 259; Conservative 0; Mismatches 275; Indels 1; Gaps 1;

```
Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              |||  ||  |  ||  |  ||| |||  ||| | |  ||| | ||  ||  ||  | ||     |
Db        488 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 547

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  ||  |  |  |  ||  |  |||     |  |  ||  |  |  ||  |  ||   |  |  |     |
Db        548 GGGGGGGGGGGGGNGGGGGGGGGGGGGGGCGGGGGGGGGGGGGGGGGGGGGGNNGGGGGGG 607

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              |  |  ||  ||     |  |  ||  |  ||  |  |||  |  ||  ||  |||    |  |  |
Db        608 GNGGGGGGGGNGGGGGGNGNGGGGGGGGNNGGGGGGGGNGGGGGGGGGGGGGGNGGGGGG 667

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              |  |  |||||  |  ||  |||  ||  ||    ||  |  ||  |||  |       |  |||  |
Db        668 GGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGCGGGGGGGGGGGGG 727

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |     |||  |  ||  ||     ||  |  |  |  ||      |||  |  |||     |  |
Db        728 GGGGGCGCGGGGGGCGGGCCCNGGGGGGCGGCGGGGGCGCGCGCGGGGCGCGCGGGGCGC 787

Qy        301 ACGCCCGAGCG-GCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGA 359
              ||  |||  |||  ||  ||||  |||        |||       |  ||  ||  ||  |  |  ||  |
Db        788 CCGGCCGGGCGCGCGGGGCGCGGGCGCGCCGGGNCGCGGGGGGGCGCCGCGGGGGCNCGG 847

Qy        360 TGAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGAT 419
              |  |  |  ||  |  ||  |  |  ||  |  |  |  ||  ||  |||||
Db        848 GCGGGCGCGCCCCGGNCCGCCCGGCGCGCCCCGCGCGNGCCGCGCCGCGCGGGCGGGCCG 907

Qy        420 CCTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGC 479
              |  ||    |||  |  |  ||    ||  |||  |  |  |  |||  |||
Db        908 CGCGGNNGGGGCGGCCGGGCGCGCGGGCGGGGCGCCGGGCCGGGGCCCGCCGGCGGCGGC 967

Qy        480 GTACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              |  |||  |||||||  |||    |  ||||  |||  |  |  |  |  |  ||  ||
Db        968 GGGCCGCCGGCGGGGGGCCCCGCGGCCGCGGGCNGCCCGGCCCGGCGCGCGGCGG 1022
```

RESULT 8
BQ939440
LOCUS BQ939440 1089 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT_8929085 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6484345 5', mRNA sequence.

```
Db        553 GGGGGGGGGGGGGGGGGGGGGNNGGGGGGGGGGNNGGGGGGGGGGGGGGGNGGGGGGGGG 612

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300

Db        613 GGGGGGGGGGGNGGGGGGNGGNNNNNGNGGGNGGGGGGGGGGGGGGGGGNGGGGGGGGGG 672

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360

Db        673 NNNNNNNGGGGGGGGGGGNNGGGNGNNGGGGGGNGGGGGGGGGGGGGNGNGNGGGGNGGNG 732

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420

Db        733 GGGGGGGGGGGGGGGGGGGGGGGGGNGGNGGGGGGGGGGGGGGNNNGGGNGGGGGGGGG 792

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480

Db        793 GNGGGNNGGGGGGGGGGGGGGGNNGGGGGGGGGNGGGGGGGGGGGGGGGNGGGGGGGGG 852

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTG 533

Db        853 GNNNNGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGNNGGGGGGGGGGGGGGG 905
```

RESULT 5
AZ200896
LOCUS AZ200896 843 bp DNA linear GSS 31-AUG-2000
DEFINITION SP_1011_A2_D05_T7A Strongylocentrotus purpuratus, purple sea urchin , sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=1011 Col=10 Row=G, DNA sequence.
ACCESSION AZ200896
VERSION AZ200896.1 GI:8395810
KEYWORDS GSS.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.
REFERENCE 1 (bases 1 to 843)
AUTHORS Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and Hood,L.
TITLE A sea urchin genome project: Sequence scan, virtual map, and additional resources
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
MEDLINE 20402566
COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 1011 row: G column: 10
Seq primer: T7
Class: BAC ends
High quality sequence stop: 843.
FEATURES Location/Qualifiers
source 1. .843
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="Plate=1011 Col=10 Row=G"
/clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli DH10B"
BASE COUNT 56 a 56 c 596 g 69 t 66 others
ORIGIN

Query Match 18.0%; Score 96; DB 17; Length 843;
Best Local Similarity 46.6%; Pred. No. 6.2e-09;
Matches 249; Conservative 0; Mismatches 285; Indels 0; Gaps 0;

```
Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60

Db        310 GGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGNGGGGG 369

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120

Db        370 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 429

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180

Db        430 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNGGGGGGNGGGGGGGGG 489

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240

Db        490 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 549

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300

Db        550 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGNGGGGGGGGGGGG 609

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360

Db        610 GGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGNGGGGGGN 669

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420

Db        670 GGGGNGGGGGGGGGNGGGGGGNGGGGGGGGGGGGGGNGNGNGGGGGGGGGGGGGGGGGG 729

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480

Db        730 GGGGGGGGGGGGGGGGNGGGNNGGGGGGGGGGGGGGGGGGGNNGGGGGGGNGNGGGGGG 789

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534

Db        790 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGNGNGGNNG 843
```

RESULT 6
BQ930965
LOCUS BQ930965 1089 bp mRNA linear EST 20-AUG-2002
DEFINITION AGENCOURT_8931577 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6483877 5', mRNA sequence.
ACCESSION BQ930965
VERSION BQ930965.1 GI:22345996
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1089)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2669 row: c column: 14
High quality sequence start: 28
High quality sequence stop: 180.
FEATURES Location/Qualifiers
source 1. .1089
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6483877"
/clone_lib="NIH_MGC_40"
/tissue_type="carcinoma, cell line"

```
            genomic clone Plate=1029 Col=7 Row=K, DNA sequence.
ACCESSION   AZ194927
VERSION     AZ194927.1  GI:8378106
KEYWORDS    GSS.
SOURCE      Strongylocentrotus purpuratus.
  ORGANISM  Strongylocentrotus purpuratus
            Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
            Echinoidea; Euechinoidea; Echinacea; Echinoida;
            Strongylocentrotidae; Strongylocentrotus.
REFERENCE   1  (bases 1 to 741)
  AUTHORS   Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R.,
            Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray
            ,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J, Davidson,E.H. and
            Hood,L.
  TITLE     A sea urchin genome project: Sequence scan, virtual map, and
            additional resources
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
  MEDLINE   20402566
COMMENT     Contact: Cameron, RA, Davidson, EH, Hood, L
            Division of Biology 156-29
            California Institute of Technology
            Pasadena California 91125, USA
            Tel: (626) 395-8421
            Fax: (626) 793-3047
            Email: acameron@caltech.edu
            Plate: 1029  row: K  column: 7
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 741.
FEATURES             Location/Qualifiers
     source          1. .741
                     /organism="Strongylocentrotus purpuratus"
                     /db_xref="taxon:7668"
                     /clone="Plate=1029 Col=7 Row=K"
                     /clone_lib="Strongylocentrotus purpuratus, purple sea
                     urchin, sperm genomic BAC library"
                     /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli
                     DH10B"
BASE COUNT       32 a     82 c    510 g     33 t     84 others
ORIGIN

  Query Match             18.1%;  Score 96.6;  DB 17;  Length 741;
  Best Local Similarity   48.8%;  Pred. No. 4.8e-09;
  Matches  261;  Conservative    0;  Mismatches  272;  Indels    2;  Gaps    1;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||| || |  | |  ||| ||| ||| | | ||| | || || ||   ||      |
Db        180 GGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNGGGGGGGGG 239

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  || | | |||| | |||    |  |  || | |  || | ||   ||  |  |
Db        240 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 299

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || | || |||   | ||| || |  ||  | |  ||||| || || | ||   |  |
Db        300 GGGGGGGGGGGGNNGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGNNGG 359

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              | |  | ||||| | || ||| || ||   ||  |  || ||| |      || ||| |
Db        360 GGGGGGGGGGGGGNGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGCGCGGGGGGG 419

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |  |  ||| |  | ||  |||| || | |  ||       | | ||   |
Db        420 GGGGGGCGGGGGGGGGGGCGGGCGGGGGGCGCGGGGGGGCCCGGGGCGGGGGGGGGGGG 479

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
                | | ||| ||||  || |    ||    || |  |  |  |   |  ||
Db        480 GGCGGGGCGGGGCGGGGCGGGGGGCGGCGGGGGGGGGGGCGCGGGGGGCGGGGGGGGGG 539

Qy        361 GAGCTGACGCGGGGTATGGAGC--TGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGA 418
              | |  | || ||||   || ||  | ||  |  || |  |      |  || |  |||
Db        540 GGGGGGGCGGGGGGCGNGGGGCGGGGGCGCGGGGCGGGGCGGGGGGCGNGGGCCGGGGCG 599

Qy        419 TCCTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGG 478
              |  ||||  || ||||| |||        | || ||      |  ||   ||  ||||
Db        600 CCGGGGGGGGGGGCGGGCGGGGGGCGGGGGGCGGGGGGGGCGGGGCNGCGGGGGCGCGNN 659

Qy        479 CGTACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTG 533
              |  | || || |||  | | | | |    ||  | ||  | || ||||||  | |
Db        660 NGGGCNGGGGGGGGGGCGGGCGGCGGGGCCGGNGGCCGGGGCCGGGGGGGCGGCG 714


RESULT 4
BQ940345
LOCUS       BQ940345                949 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_8806832 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:6338352
            5', mRNA sequence.
ACCESSION   BQ940345
VERSION     BQ940345.1  GI:22355823
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 949)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2531  row: h  column: 01
            High quality sequence stop: 271.
FEATURES             Location/Qualifiers
     source          1. .949
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6338352"
                     /clone_lib="NIH_MGC_47"
                     /tissue_type="neuroblastoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT       55 a     62 c    563 g     86 t    183 others
ORIGIN

  Query Match             18.0%;  Score 96.2;  DB 14;  Length 949;
  Best Local Similarity   42.6%;  Pred. No. 5.7e-09;
  Matches  227;  Conservative    0;  Mismatches  306;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              |||          || |  ||| |||  ||| |    |||   || ||  | | ||     |
Db        373 GGGNNNNGNNNGGGGGGGGGGGGGGGGGGGGGNNGGGGGNNGGGGGNNGGGGGGGGGGGG 432

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |       | | |  | | |||    |  |  || |    ||   ||   ||  |    |
Db        433 GNNNNNNNGNGGGNNGGGNGGGGGGGGGGGGGGGGGGNNNGGGNNGGGGGGGGGGGGGG 492

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || | || |||     ||| ||    ||  | |  ||||| || || | ||   | | |
Db        493 GGGGGGGGGGGGGGGNNGGGGGGNNGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGNGGG 552

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
```

```
                     /db_xref="taxon:169015"
                     /clone="hasp001xa01f"
                     /clone_lib="Heterobasidion annosum - Scots pine infection
                     stage subtraction cDNA library (hasp)"
                     /dev_stage="Seedling roots of scots pine were infected for
                     6 days with H. annosum"
                     /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive
                     hybridization cDNA library was constructed from  scots
                     pine roots infected for 6-days with mycelia of
                     Heterobasidon  annosum (FP5)."
BASE COUNT       20 a      28 c    1036 g      28 t      85 others
ORIGIN

  Query Match             18.4%;  Score 98;  DB 13;  Length 1197;
  Best Local Similarity   46.4%;  Pred. No. 2.6e-09;
  Matches  248;  Conservative    0;  Mismatches  286;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||| || |  || |  ||  ||| ||| | |  ||||| || ||  ||    |
Db        487 GGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGTGTGGGGGGGGNNGGGGNGGNG 546

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  ||||   |||||  | |||    |  | || | |   | | ||   ||  |    |
Db        547 GGGGGTGGTGGTGGGTGGTGGGGGNGGNGGGGGGGGGGGGNGGGGGGGGGGGGGGGGNGG 606

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || | || |||   | ||| || |  ||  | |  ||||| || || | ||   | | |
Db        607 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 666

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              | |  | |||| | || ||  || ||   ||  |  || ||       |  ||| |
Db        667 GGGGGGGGGGGGGGGNGGGGGNNGGGGGGGGGGNNGGGGGNGGNNNGGGGGGGGGGGGN 726

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              | | | | |    | |  | || ||| | |  ||  | | | |    |
Db        727 NGGGGGGGGNGNNNGGGGGGGGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGNGGGGGG 786

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |   | | ||  |||   || |   | ||    || ||  |  |       |  ||
Db        787 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGNNGGGGGGGGGGGGNGGGGGGNGGGGN 846

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              | | |  | |||  || |  |  || |   | |  |          ||  | ||||
Db        847 GGGGGGGGGGNGGGGGGGGGGGNGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGG 906

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              ||||  ||  |||  |||     |  | | ||     |  |    ||  | |  || | |
Db        907 GGGGGGGGGGGTGGGGGGGGGGGGTGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 966

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              || || ||| || |||    | |||||  |      || | |||||  | ||
Db        967 NGGNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGTTGGGGGGGGGGGGGGGGG 1020


RESULT 2
BM911600
LOCUS       BM911600               1478 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6614915 NIH_MGC_41 Homo sapiens cDNA clone IMAGE:5466410
            5', mRNA sequence.
ACCESSION   BM911600
VERSION     BM911600.1  GI:19361979
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1478)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1966  row: m  column: 03
            High quality sequence stop: 272.
FEATURES             Location/Qualifiers
     source          1. .1478
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5466410"
                     /clone_lib="NIH_MGC_41"
                     /tissue_type="amelanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
BASE COUNT      135 a    253 c    815 g     76 t    199 others
ORIGIN

  Query Match             18.2%;  Score 97.4;  DB 14;  Length 1478;
  Best Local Similarity   42.6%;  Pred. No. 3.4e-09;
  Matches  227;  Conservative    0;  Mismatches  306;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||| || |  || |  ||| ||| ||| | |  ||| | ||    || | ||     |
Db        499 GGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGNNNNGGGGNGGNNNGGG 558

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  || | | |||| | |||     |     || | |   | | ||   ||  |
Db        559 GGGGGGGGGGGGGGGGGGGGGGNGNGGGNNNGGGGGGGGGNGGGGGGGGGGGGGGGGGN 618

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || | || |||   | ||| | |  ||  | |  ||||| || || | ||   | | |
Db        619 GGGGGGGGGGGGGGGGGGGGGNGGGGGGGGGNGGGGGGGGGGGGGGGGGGGGGGGGGGG 678

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              | |  | |||| | |  ||| || ||   ||  |  || ||| |      | |||
Db        679 GGGGNGGGGGGGGGGGGNGGGGGGGGGGGGGGGGGGGNGGGGGGGGGGGGNGGGGGGNG 738

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              | |  |||  | | |  || ||  |  |  |  |  | | |      |
Db        739 GGGGGGGGGGGGGGGGNGGGGGGGGGGGGGNNGGGGGGGGNGGGGGGNGGGGGGGGGG 798

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |   | | ||  |  | || |  | |    || ||  |  |       |  ||
Db        799 GGGGGGGGGGGGGGGNGGGGGGNGGGGGGGNGGGGGGGGGNNGGGGGGGGGNGGGNGGGG 858

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              | |  |  | |||    |  |  | || |  | |  |              | ||||
Db        859 GGGGGGGGGGGGGNGGGGNNGGGGNGGGGGNGGGGGGGGGNGGNNNGNNGGGNGGGGNNG 918

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              |     | |||  | |          | ||      |||    |  |  || | |
Db        919 GNNGNNGGNGNGGGGNGGNGGNGGNNGNGNGGGGGGGGGCCCGNNNNNGGNGNNGGGGNG 978

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTG 533
              | || ||| || |||   | ||  | |  | || | ||| |  | |
Db        979 NNNNNGGGGGGGGGGNGGGGNNGGGGGNNGNGNNGNNGGNGGGGGNGNGGG 1031


RESULT 3
AZ194927
LOCUS       AZ194927               741 bp    DNA     linear   GSS 30-AUG-2000
DEFINITION  SP_1029_A1_F04_SP6E Strongylocentrotus purpuratus, purple sea
            urchin, sperm genomic BAC library Strongylocentrotus purpuratus
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 684.535 Seconds
(without alignments)
12633.968 Million cell updates/sec

Title: US-09-914-286-1_COPY_10609_11142
Perfect score: 534
Sequence: 1 gggtcggtgttggtgacggg.....cgctggcgtgggggctgtgg 534

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	98	18.4	1197	13	BI416470	BI416470 hasp001xa
2	97.4	18.2	1478	14	BM911600	BM911600 AGENCOURT
3	96.6	18.1	741	17	AZ194927	AZ194927 SP_1029_A
4	96.2	18.0	949	14	BQ940345	BQ940345 AGENCOURT
5	96	18.0	843	17	AZ200896	AZ200896 SP_1011_A
6	95.8	17.9	1089	14	BQ930965	BQ930965 AGENCOURT
7	95.6	17.9	1136	14	BQ943816	BQ943816 AGENCOURT
8	95.4	17.9	1089	14	BQ939440	BQ939440 AGENCOURT
9	95.2	17.8	855	9	AL572700	AL572700 AL572700
10	94.8	17.8	937	14	BQ941425	BQ941425 AGENCOURT
11	94.6	17.7	993	14	BQ958704	BQ958704 AGENCOURT
c 12	93.8	17.6	1089	13	BM468944	BM468944 AGENCOURT
13	93.8	17.6	1184	14	BM911719	BM911719 AGENCOURT
14	93.4	17.5	888	17	AZ189327	AZ189327 SP_1014_A
15	93.4	17.5	897	13	BI416517	BI416517 hasp001xd
c 16	93.4	17.5	1362	13	BM556092	BM556092 AGENCOURT
17	93.2	17.5	748	17	AZ187748	AZ187748 SP_1009_B
18	93	17.4	1243	14	BM913931	BM913931 AGENCOURT
c 19	92.8	17.4	877	17	AQ788449	AQ788449 HS_3130_A
20	92.8	17.4	931	13	BI416575	BI416575 hasp001xg
21	92.8	17.4	1090	17	AG128386	AG128386 Pan trogl
22	92.4	17.3	863	13	BI416699	BI416699 hasp001xp
23	92.4	17.3	905	13	BI416651	BI416651 hasp001xm
24	92.4	17.3	939	14	BQ890606	BQ890606 AGENCOURT
c 25	92.4	17.3	1006	10	BE034822	BE034822 ML04H02 M
26	92.2	17.3	715	9	AL570062	AL570062 AL570062
27	92.2	17.3	943	13	BM415351	BM415351 OP20425 M
28	92.2	17.3	953	14	BQ948564	BQ948564 AGENCOURT
29	92.2	17.3	1047	14	BQ894173	BQ894173 AGENCOURT
c 30	92	17.2	729	10	AW583629	AW583629 ia02b12.y
c 31	92	17.2	795	17	AZ188474	AZ188474 SP_1012_A
32	92	17.2	1011	17	AG062256	AG062256 Pan trogl
c 33	91.8	17.2	915	17	AZ195437	AZ195437 SP_1030_A
c 34	91.8	17.2	932	17	CNS06KYG	AL403502 T7 end of
35	91.6	17.2	782	12	BG809582	BG809582 mgct001xa
36	91.6	17.2	798	13	BI416695	BI416695 hasp001xo
37	91.6	17.2	864	17	AG060559	AG060559 Pan trogl
38	91.6	17.2	918	17	AQ895329	AQ895329 HS_4832_A
39	91.4	17.1	542	12	BG085788	BG085788 H3117F03-
c 40	91.4	17.1	610	9	AL514195	AL514195 AL514195
c 41	91.4	17.1	737	17	AZ193539	AZ193539 SP_1023_B
42	91.4	17.1	817	17	AZ196702	AZ196702 SP_1032_B
43	91.4	17.1	993	17	AG137966	AG137966 Pan trogl
44	91.4	17.1	1015	13	BI416650	BI416650 hasp001xm
45	91.4	17.1	1044	17	AG126158	AG126158 Pan trogl

ALIGNMENTS

RESULT 1
BI416470
LOCUS BI416470 1197 bp mRNA linear EST 15-AUG-2001
DEFINITION hasp001xa01f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xa01f, mRNA sequence.
ACCESSION BI416470
VERSION BI416470.1 GI:15187493
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 1197)
AUTHORS Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026,S-750 07, Uppsala, Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES Location/Qualifiers
source 1. .1197
/organism="Pinus sylvestris/Heterobasidion annosum"

THIS PAGE BLANK (USPTO)

```
                     QSGAMTDGLVEADRARMARSGVLPLPPEEGLRLFDAALASDEAVLAPVRIDTGALRAG
                     EAPPVLRALVPAAAARRTARAAAAPSASSLAERLEEMPEAEREKTVLDLVRAEVAAGL
                     GHASDRTVRPEHAFQDLGFDSLTAVELRNRLNKATGLRLPATLVFDHPTPALLARHVF
                     MAAAGAAEPALVDSLLADLDRVEQELVTTLAASEARDRILSKLQGLLAIAGEFGKAPA
                     QEGPGTGGSDLESATDDEVFDLLGKEFGIS"
     gene            21273. .31847
                     /gene="pteA2"
     CDS             21273. .31847
                     /gene="pteA2"
                     /codon_start=1
                     /transl_table=11
                     /product="modular polyketide synthase"
                     /protein_id="BAB69304.1"
                     /db_xref="GI:15824140"
                     /translation="MSCRYPGGVDSPEDLWELVAGGRDAVSEFPADRGWDLDALYDPD
                     PESRGTSYAREGAFLRDLARFDAGLFGISPREALAMDPQQRLLLETSWELFERAGIDA
                     ASLAGSRTGVFAGVMNHEFLAALQNSPQDLEGYLGTGTSGSVASGRVAYTFGLEGPAV
                     TVDTACSSSLVALHLAVQSLRNGECSLAVAGGVTAMVGPATFVGFSRQRGLAPDGRCK
                     PFAAGADGTGWGEGAGLLLVERLSDARRNGHEVLAVVRGSAVNQDGASNGLTAPNGPS
                     QQRVIRQALAAAQLTTAQVDAVEAHGTGTRLGDPIEAQALLATYGQGRPDDKPLWLGS
                     IKSNIGHTQAAAGVAGIIKMVMAMRHGELPRTLHVDAPSVHVDWTAGAVRLLTERTPW
                     SASDQPRRAGISSFGISGTNAHTIIEEAPSAEAFAQDTGAPEGPEEGAESAEPVASEA
                     ARPAAGGPVPWVISGRTEEALRAQAERLLSRTDAEVPPLDVAFSLATTRTALEHRAVL
                     QGETPAELAEGLRAVASGSPARGVARGLSGTGGRVGFLFSGQGSQRLGMGRELYASYP
                     AFAAAYDEVCAQLDAVVDVDSEELHRTGAAQPALFAVEVALFRLLESWGIRADYVAGH
                     SVGELAAAHVAGVLSLDDAAKLVSARARLMQALPAGGAMVAVQATEDEVLPHLTDAVG
                     IAAINGPRSVVVSGAEDAVTAIAEAFRQQGRKSSRLKVSHAFHSPLMDPILEEFAEVV
                     RSLTFHRPQIPVVSNLTGRLAEPYTPQYWVRHVREAVRFADGVRTLNDLGVTTFVEIG
                     PGGVLSALAEGCLDDDVVTVPALRADRPERVALVAAVSELHAHGVSPDWHAFFPGARR
                     VDLPTYAFQYERYWLDASPAAPGGVRAAGLGSADHPLLGAAVSLAGGDERLLTGRLSL
                     RTHPWLSDHAVLGTALLPGTAFVELAVRAADEAGCALLEDLTLEAPLVVPEQGGVAVQ
                     VWVGAVDGTGRRPLTVHSRPEDDSDLPWVRHATGFVTEGSEEAGAAAADRTLTAWPPT
```

Query Match 32.9%; Score 85; DB 1; Length 78210;
Best Local Similarity 64.2%; Pred. No. 5.2e-05;
Matches 165; Conservative 0; Mismatches 80; Indels 12; Gaps 2;

```
Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||| |||||||||| | |    | |           ||| ||||||| |||||| | |||
Db      60220 CTGGCCGCGCTGTCGCACCAAGACCGCGCACGCGCCCTCGTGGACCTCGTGCGCACGCAC 60279

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
               ||   || ||  ||   | ||  |||     |     |  || | | ||  ||||||| ||
Db      60280 GCCGCCGCCGTCCTCGGCCACCGCGGGGCGGGCGCCGTCGAGGCCGA--CCGGCCCTTCC 60337

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              | |          || |||||||||||||| | |  |  || ||||| ||||||||  ||
Db      60338 GGG----------ACATCGGCTTCGACTCCCTGACCGCGGTGGAACTCCGCAACCGGATC 60387

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              | ||||||||| |||||     |||||| |||||| |||||||||||| ||||| | ||
Db      60388 AACAAGGCCACCGGCCTCTCCCTGCCCGCGACGCTGATCTTCGACCACCCCACGGCCGCG 60447

Qy        241 GCGGTCGCCGCGCGCCT 257
              |   ||||||| | |||
Db      60448 GACCTCGCCGCCCACCT 60464
```

Search completed: June 17, 2003, 17:55:21
Job time : 503.923 secs

```
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES              Location/Qualifiers
     source           1. .78210
                      /organism="Streptomyces avermitilis"
                      /db_xref="taxon:33903"
                      /note="Actinobacteridae"
     CDS              complement(1. .1254)
                      /codon_start=1
                      /transl_table=11
                      /product="putative lipoprotein"
                      /protein_id="BAB69301.1"
                      /db_xref="GI:15824137"
                      /translation="MGEALITLCTRSTRLAATAGVLGGVLAVTALGGTSNAATSGNGA
                      GNHPTSQARINQAGAQEASDARIEITPGQGTYSVGTNHPVSVTVTSGKLTKVTMTAVA
                      TGAEIPGTLSTDGTSWKPNGRLERATKYQIAAEAGDVKGRSTTGSATITTVSPANDFI
                      GHLSPEDGSTVGVAMPVTVNFDKAISDKAAVQSEIRVSSSSGQRVVGHWFNDHRVDFR
                      PENYWRPGSTVTVTLNRHGIQKTVTFKIGRSQISTVDAKTKQMTVVRDGKTIKTIPIS
                      SGSPEHPTYNGQMVISEKFRQIHMNGASVGLTEKNGKPSYDIKAVPHAMRLTDSGTFI
                      HGNYWGTDSVFGEANTSHGCVGLKDVRGDSHGRQPAAWFFDHSIVGDVVIVKNSEEKT
                      MLAPDNGLSDWNMSWSEWVAGSPAT"
     CDS              complement(1256. .2290)
                      /codon_start=1
                      /transl_table=11
                      /product="putative transposase"
                      /protein_id="BAB69302.1"
                      /db_xref="GI:15824138"
                      /translation="MPVEFLSHDQVSRYGRFASAPTPGELEQFFRLDAKALESARSKR
                      APANRLGWAVQWGCVRMLGVFPTEDLSVVPEVVVRFAAEQLGVDPGEFAAYGARRQNR
                      YEHAWEIRDAYAYRESPAAEAQVREFLATRVWASWEGPRALFDRAVGWLVDNRVLLPG
                      ITTLTRLVAEVRAQENAALYRTLDEAVPEKLRQSMRDLLKVPEGRRVSELERLRRSPM
                      GVSGSADRLAFSERIRMEAVAMFTAGPGSTDIAKELRVSVQSVQRWRSGLAGDRPGRG
                      AVPQPVSVELSETLFAALDEERAKGFAPSWTIRYLHHDFPIFGQRYAYKWSQGCLHSF
                      GRSAQGYRDQ"
     gene             2935. .21081
                      /gene="pteA1"
     CDS              2935. .21081
                      /gene="pteA1"
                      /codon_start=1
                      /transl_table=11
                      /product="modular polyketide synthase"
                      /protein_id="BAB69303.1"
                      /db_xref="GI:15824139"
                      /translation="MVESELNATGNRDCSASADSADGIAVIGLACRLPGAADPAAFWR
                      LLSEGADAITDVPPDRWDGAVVADASAPDRTGIRRGGFLDSIGDFDAGFFGLSPKEAA
                      AMDPQQRLVLELAWEALEDAHVRAATLRSTRTGVFVGAMWDDYATLHHRSGLTAISPH
                      TVTGLHRSIIANRVSYFLGLNGPSLTVDSGQSSSLVSVHLACESLRKGESTIALAGGV
                      NLNIVPESALGAAKFGGLSPDGRCFTFDARANGYVRGEGGGIVVLKPLARAVADGDPV
                      YCVIRGSAVNNDGGGDGLTVPLQSGQEQVLRLAYERAGVNPADVGYVELHGTGTKVGD
                      PIEAAALGAVLGTGREPGRPLRVGSAKTNVGHLEGAAGITGLLKAVLSLNHRELPASL
                      NFGTPNPAIPLDRLNLRVQTEHSEWAAEDGRPLLAGVSSFGMGGTNCHIVLASGAAVE
                      AGEGATAGGGAAPTFWPLSAKTASGLRAQAAALLAHTEAPLGLALPDVGWSLATGRTA
                      FEHRAVVVGEVRDDFLRALRELSTGGIDAALTTGRAGPRGELAFLFSGQGSQRAGMGR
                      ELAESVPAFASALDEVCAHLDPLLPRPLREVMFAPVGTGEAAELDQTLYTQTSLFAVE
                      VALFRLLESWDITPDVLMGHSIGELAAAHVAGVLSLADACALVAARGRLMQALPAGGA
                      MIAVQAAEDEVRARIGERTDRVSIAALNGPSTVVSGDEDLATEIADAFAALGRKTSR
                      LRVSHAFHSPHMEPMLAEFRNVAEGLTFHAPRIPIVSNVTGRLSEEPSGYEGCTAAYW
                      VRHVREAVRFADGVARLDEQGVRTYLELGPGGVLASMARAGADSESLFVPALRTRRPE
                      AQALLAAVASLHVHGVEPDWDALFGGGRGARRKVALPTYAFERQRYWLDGDTTSPVAT
                      TGWAPVAPQPDAIAETDSAPVGALGRRLAGLAESERDEAVLDMVRAHIAAVLGHAEAR
                      QVETEWTFKDLGFDSLSSVELRNQLSEATGLRLPSGLLFDHPTPAALARHLAAQSLGG
                      GDAAAAAPATVADPHEPIAIVAMSCRLPGGVSSPEDLWRLLASGGDGISGFPTDRGWD
                      VETLYDPEPGTPGKSSTRHGGFLHEAAEFDAGFFGISPREAAAIDPQQRLVLETAWEA
                      FERAGIDPATLRGSRAGVFIGATAQDYGPRLHEPADGVEGYLLTGTTTSVASGRVAYS
                      FGLEGPAVTVDTACSSSLVALHLAVQSLRSGECTMALAGGVTVMSTPGMFVEFSRQRG
                      LSADGRCKAFAAGADGTGWAEGVGMLLVERLSDARRNGHQVLAVVRGSAVNQDGASNG
                      LTAPNGPSQQRVVRQALAAAGLAPADVDAVEAHGTGTLLGDPIEAEALIAAYGQGRDG
                      DRPLWLGSLKSNIGHTQAAAGVAGVIKMVMAMRHGVLPTRTLHVDEPTPHVDWSASGVR
                      LLTEAVEWPARDLPRRAAVSSFGISGTNAHVIVEQAATSEEPARGTAVAGPVPWVVSA
                      KSDVALREQAGRLLSFLDAEGVGGAGVRPEDVALSLATSRAVLERRAAVVGGDLAELR
                      QGLEALASGASGAAPVGGAVGGRVGFLFSGQGSQRIGMGRELYAAYPVFAAAYDEVCA
                      FLDAPVDVDSEELNQTGTTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVA
                      GVLSLEDAAKLVSARAALMQALPDGGAMVAVQATEDEVLRHLTDGVGIAAVNGPRSVV
                      VSGAEDAVTAVGEIFAGRGRKTSRLKVSHAFHSPLMDPMLEEFAAVVRGLTFSESRIP
                      VVSNLTGRLAEVYTPEYWVRHVREAVRFADGIETLGELGVTTFVEVGPGGVLSALAQG
                      CFDDGTDIVTVPVLRADRPEPYALVTAIGRLHTHGISPDWQVLFPGARRVDLPTYAFQ
                      RERYWLDAPRAATAGAPDAMDAEFWDSVESEDRASLGALLGLEPTELDVVASKLSAWR
                      RKRREQSTTDGWRYRITWQPLDDMVRAAPSGTWLYVVSEETAWTEAIRAGLTELGVAL
                      VPFAVGEDTDRDALARAFADAGQEQADGILFAAAQGEAGANALQRLVLLVQALGDAGI
                      EAPLWCLTSGAVSTGTADLLTDPVSAQLWGLGRIVALEQPQRWGGLVDLPAELDSRAL
                      ERLAGLLAQADEDQLAVRASGVYGRRLVRAPLSAAPADAPWSPRGTVLVTGGTGALGG
                      QVARWLAGAGAEHLVLTSRRGPDAPGAAELKAELEELGAQVTVVACDAADRDALAELF
                      GRYIVNAVVHTAGVLDDGLVESLTPERLNHVLRPKVDAALHLHELTRDRQDLDAFVLF
                      SSMTGVWGNGGQGAYGAANAFLDALAEHRRAQGVPALAVAWGSWADGGMADGAAGDHL
                      RRRGVRAIAALPAISVLHGALTHGETSMTVADVDWERFVPAFVGTRPCPLLLGVPEAR
                      EALEARARAAQSPEVPVSALVQRLLGATPGEQGRILLDLVREQAATVLGHTGKGAFEA
                      DRTFRETGFDSLTAVELRHRLNTATGLRLPTTLVFDHPTPTALARHLGEELLGRHGLE
                      TAAAQATAAAAADEPLAIVGMSCRLPGGVQSPEDLWELLSSGRDAMSGFPTDRGWDID
                      ALYDPDPDRPGKTYSREGGFIQGADRFDAGLFGISPREALAMDPQQRLLLETAWEAFE
                      RAGIALASVRASRTGVFIGTNGQDYANGLRNAPEEIEGYALTGKAASVVSGRISYTFG
                      LEGPAVTVDTACSSSLVALHLAAQALRSGECSMALVGGVTVMTTPDLFVEFSRQRGLS
                      PDGRCKAFAAGADGTGWGEGVGLLLVERLSDARRNGHQVLALVRGSAVNQDGASNGLT
                      APNGPSQQRVIRQALASAGLTPADVDAVEAHGTGTRLGDPIEAQALLATYGQEHSDDR
                      PVWLGSVKSNIGHTQAAAGVAGVIKMVLAMRHGVLPQTLHVDEPTAHVDWSAGAVRLL
                      TEPVQWPGTTGPRRAAVSSFGIGGTNAHTIIEEAPAAAAAEPAREHRPVPVPWVLSAK
                      SEVALRAQAERLLAFATDDVSPVDAGFSSATTRSALEHRAAVIGSDPAELRAGLEALA
                      AGEPAANVVAGRAHSADKVGFLFSGQGSQRLGMGRELYAAYPVFAAAYDEVCALLDAT
                      VDVDAETLHQTGSTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVAGVLSL
                      EDAAKLVSARAALMKALPAGGVMVAVQAAEDEMLPHLIDRAGIAAINGPRSVVISGAE
                      DAVTAIAEVFAAQGRKTSRLKVSHAFHSPLMDPMLEEFAEVVRGLTFSEPWIPVVSNL
                      TGRLTESYTPEYWVRHVREAVRFADGVLTLHELGVTTFVEIGPGGVLSALVQGCLDDG
                      TDIVTIPALRADRPEPQALTTAYAQLHVSGAEVDWHAFFPGARRVELPTYAFQRERYW
                      LDAPASAGDMRAVGQGEAGHPLLGAAVPLADSVGHLLTGRLSKHTHPWLVDHAVSGVA
                      LLPGTAFVELAITAADAVGCDLLEELTLETPLLVPERGGVALQVHVGADDGSGRRPLT
                      VHSQIDDGDAYAHEDDSERAWVRHASGVLTVATGGPVDGSLTAWPPVGAEPVDVDGFY
                      DRLAGMALDYGPVFQGLRAAWRSGDDFFAEVELPGAERTDAGAYGLHPALFDAALHTV
                      WLGAVEPEAGTGNGLLPFAWSGVRLAAAGASDLRVKVSRAGTGTSTVSLVLADGTGGP
                      VAEVDSLTLRPVPADQLRAGSGSDDAVFGLKWTPVSLPSAPDGMRIETYEDLAALRAT
                      DPPADTEADAAAAADAVVVPCPTGTGADLATRVREVTHAVLELVQWWLAEERPGRLVL
                      VTRGGDLAQSAVCGLVRSAQTENPDRIVLVETSSDAHAGAGSGVDDADLARVLPGLIA
                      SGEPQAAVHADDEVRVPRLVRVAVPDAEPELPGLGSGTVLLTGASGGLGGLFARHLVA
                      EHGVRSLLLVSRRGGDAPGAAELTADLTAQGADVTWAACDVADRAAVRALLAGPGQSL
                      SAIIHTAGVLDDGIIGSLTPERLDAVFRPKVDAALNLHDLAGELSGDLSAFVLFSSVA
                      GTLGTPGQANYAAANTFLDALAEHRRAAGQRATSLAWGLWAQDSESAMTGSLDHTDLT
                      RIQRMGLAAIPSSDGLRLFDAALALDRAVVAPLRLDLGGLQGTVPPVLRGLRGTVRGT
                      ARRTAQTAPKGSLGQRLAGLAEAERERQVVLDLVRTEVAAVLGHAGAQSVGADDSFKDI
                      GFDSLTAVELRNRLNSAVGMRLPATLIFDYPNPLTLAHFLTGEAAGGEAATAVPTTLS
                      TATGTADEPIAIVGMACRYPGGVQSPEDLWQLVFSGRDAVSGFPEGRGWDVEQLYDPN
                      PDQWGTSYTREGGFLHDAADFDAEFFGISPREALAMDPQQRLLLETSWEAFERAGIDP
                      ATVRGSRTGVFAGVMYHDYGGRVHTSPPGLEGYLVNGSAGSIASGRVSYTFGLEGPAV
                      TVDTACSSSLVALHLAAQALRSGECSMALVGGVTVMAGPSVFVEFSRQRGLSADGRCK
                      AFGAGADGTGWAEGVGMLLVERLSDARRNGHKVLAVVRGTAVNQDGASNGLTAPNGPS
                      QQRVIRQALANAGLSSDEVDAVEAHGTGTRLGDPIEAQALIATYGQERSEDRPLWLGS
                      LKSNIGHSQAAAGVGGVIKMVMAMQHGVLPQTLHVDEPTPHVDWSAGAVQLLTEAVEW
                      PASDGPRRAAVSSFGVSGTNAHVIVEQAATSEVTARGTAVAGPVPWVLSAKSDAALRE
                      QAGRLLSFLDAEGVGGAGVRPEDVAYSLATSRAVLECRAAVVGEELGEFRQGLEGIVS
                      GAAPVGGAVGGRVGFLFSGQGSQRIGMGRELYAAYPVFAAAYDEVCALLDATVDVDAE
                      TLHQTGSTQPALFAVEVALFRLLESWGIRPDYVAGHSVGEIAAAHVAGVLSLEDAAKL
                      VSARAALMQALPAGGAMVAIQATEAEVLPHLTHGVGLAAINGPRSVVLSGSEDAVLDI
                      AEVFAQQGRKTSRLKVSHAFHSPLMDPMLEEFAEVVRGLTFSEPRIPVVSNLTGRLAE
                      VYIPEYWVRHVREAVRFADGLHTLHELGVTMFVEIGPGGVLSALVQGCLDDGTHIVTV
                      PALRADRPEPYALVTAVGQLHTHGISPEWQALFPGARRVDLPTYAFQRERYWLDAPQA
                      SGDVRAAGLGAADHPLLGAAIATADSDGAVLTGLLSLGTHPWLADHAVNGIVLLPGTA
                      FVELAIRAGDETGCGWIEDLTLELPLVVPERGGVTLQVAVGPEDESGRRQLSVHSRVD
                      DRSWLRHATGVLSSAEAPSAAKLGAWPPAGAEAVDLTAFYEGMAAAGLDYGPVFQGLR
                      SVWRDGDDVFAEVELPDESAAEAAAFALHPALLDAALHALAAGGLVSLDDSPALPFAW
                      SGVVVQAAGAAMVRVKLSRTATGSVTLAVADGAGDPVASVESLSLRAASAEQLPGAVG
                      GDQLFGLEWTRFTLPSATDALMIETVADFDALRESEGTPDVVVVPVVPCPAGSGRETA
                      DRVHSTTSEVLALVQWWLAEERAGRLALVTRPGDLAHAAVWGLVRSAQSENPDRIVLV
                      EAEETDEAVRVLPAALASGEPQFAVRDGEVFVPRLVKATRSATGTATGTPDFGVGPVL
                      LTGASGALGGLVARHLVAEHGVRSLLLLSRRGAEAPGAVELEAELAAWGAEVRWAACD
                      VADREAVSGMLNGLGEQSLSAVVHTAGVLGDGIVASLTPERMREVFRPKVDAVLNLHE
                      CTRDMGLAAFVVFSSVAGMVGSAGQASYAAANSFLDAFSAHRRREGLPAISLAWGVWE
```

```
  Query Match            35.0%;  Score 90.2;  DB 1;  Length 90445;
  Best Local Similarity  63.0%;  Pred. No. 7.5e-06;
  Matches  162;  Conservative    0;  Mismatches   83;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||| ||| |||  ||||  || | ||||    ||||||||||||||||||| || || ||
Db      27778 CTGGCCGGGCTCGCCCCGGCGGACCAGCTCACGCTGCTCCTGGACCTGGTCCGGGCGCAG 27837

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
                | ||| |||||     ||| ||              ||||   |  | | ||
Db      27838 GTCGCGGCCGTGCTCGGGCACGCCGA-----------CGCGAGCGCCGTCCGCGTCGAC 27885

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              ||   |||     ||  |||||||||||  | |   |  ||||| ||||||||||| |
Db      27886 ACGGCCTTCAAGGACGCCGGCTTCGACTCGCTGACCGCGGTCGAGCTGCGCAACCGCATG 27945

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              | |  ||||| ||||||  | |||||| |||||| |||||||| || | |  ||
Db      27946 CGGACCGCCACCGGCCTGAAGCTGCCCGCGACGCTCGTCTTCGACTACCCGAACCCCCAG 28005

Qy        241 GCGGTCGCCGCGCGCCT 257
              ||| |||||  || |||
Db      28006 GCGCTCGCCCGGCACCT 28022


RESULT 13
AX250263/c
LOCUS       AX250263               30000 bp    DNA     linear   PAT 05-OCT-2001
DEFINITION  Sequence 3 from Patent WO0168867.
ACCESSION   AX250263
VERSION     AX250263.1  GI:15984065
KEYWORDS    .
SOURCE      Streptomyces cinnamonensis.
  ORGANISM  Streptomyces cinnamonensis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 30000)
  AUTHORS   Leadlay,P.F., Staunton,J. and Oliynyk,M.C.
  TITLE     Polyketides and their synthesis
  JOURNAL   Patent: WO 0168867-A 3 20-SEP-2001;
            Biotica Technology Limited (GB)
FEATURES             Location/Qualifiers
     source          1. .30000
                     /organism="Streptomyces cinnamonensis"
                     /db_xref="taxon:1900"
                     /note="Nucleotides 60001 - 90000 of the monensin
                     biosynthetic gene cluster"
BASE COUNT     4041 a  10632 c  11060 g   4267 t
ORIGIN

  Query Match            33.7%;  Score 87;  DB 6;  Length 30000;
  Best Local Similarity  62.3%;  Pred. No. 3.1e-05;
  Matches  160;  Conservative    0;  Mismatches   85;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||  |||| |||||  | |  ||| | ||  |  ||| || |||||||| || ||||
Db      18070 CTCGGCGCGATGTCCGACCCCGAGCGGGAGGAGGCGCTGCTCGACCTGGTGTGCACCCAT 18011

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              | |  ||| ||  ||    | ||||||              | ||  ||   | |  | |  ||
Db      18010 ATCGCGGCCGTCCTCGGCTACGACG-----------CGGCCACGCCCGTCAACGCGACG 17963

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                | |  |     | ||||||||||||||| | |   |  || || ||||||||||| |||
Db      17962 CAGGGGCTGCGGGAACTCGGCTTCGACTCCCTGACCGCGGTGGAGCTGCGCAACCGGCTC 17903

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
                || ||||| |||  | ||||||  ||| || ||||||||||  |||  |||||
Db      17902 TCGGCCGCGACGGGGCTGAAGCTGCCCGCCACGTTCGTCTTCGACCATCCCAACCCGGCG 17843

Qy        241 GCGGTCGCCGCGCGCCT 257
              | | ||||||||||  ||
Db      17842 GAGCTCGCCGCGCAGCT 17826


RESULT 14
A69718
LOCUS       A69718                 5676 bp    DNA     linear   PAT 07-MAY-1999
DEFINITION  Sequence 1 from Patent WO9807868.
ACCESSION   A69718
VERSION     A69718.1  GI:4774327
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 5676)
  AUTHORS   Schupp,T., Toupet,C. and Engel,N.
  TITLE     RIFAMYCIN BIOSYNTHESIS GENE CLUSTER
  JOURNAL   Patent: WO 9807868-A 1 26-FEB-1998;
            CIBA GEIGY AG (CH)
FEATURES             Location/Qualifiers
     source          1. .5676
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT      674 a   1990 c   2205 g    807 t
ORIGIN

  Query Match            33.1%;  Score 85.4;  DB 6;  Length 5676;
  Best Local Similarity  61.9%;  Pred. No. 8.3e-05;
  Matches  159;  Conservative    0;  Mismatches   86;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||  ||| |||  | ||  || ||||| || ||||||| || ||| | || ||| | ||
Db       3072 CTCGCCGGGCTCGCGCCGGCGGAGCAGGAGGCGCTGCTGCTCGACGTCGTCCGCACGCAG 3131

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
                | |  ||||||     ||| |||    |||  |  |    ||||  | |
Db       3132 GTCGCGCTGGTGCTCGGGCACGCCGGGCCGGAGGCCGTCCGCGCGGACACGGCG------ 3185

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                      |||     ||  |||||||||||  | | |    || ||||||||||||| ||
Db       3186 ------TTCAAGGACACCGGCTTCGACTCGCTGACGTCGGTGGAACTGCGCAACCGGCTG 3239

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              || |||| |  || |||  | ||||||  |||||||  ||||||||  || | |||||||
Db       3240 CGCGAGGCGAGCGGGCTGAAGCTGCCCGCGACGCTCGTCTTCGACTACCCGACGCCGGTC 3299

Qy        241 GCGGTCGCCGCGCGCCT 257
              ||| | |||     |||
Db       3300 GCGCTGGCCCGCTACCT 3316


RESULT 15
AB070949
LOCUS       AB070949               78210 bp    DNA     linear   BCT 11-OCT-2001
DEFINITION  Streptomyces avermitilis polyene macrolide biosynthetic gene
            cluster.
ACCESSION   AB070949
VERSION     AB070949.1  GI:15824136
KEYWORDS    .
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
  TITLE     Genome sequence of an industrial microorganism Streptomyces
            avermitilis: deducing the ability of producing secondary
            metabolites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
  MEDLINE   21477403
REFERENCE   2  (bases 1 to 78210)
  AUTHORS   Ikeda,H.
  TITLE     Direct Submission
```

```
                /organism="Amycolatopsis mediterranei"
                /strain="S699"
                /db_xref="taxon:33910"
     CDS        423. .1451
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAC01706.1"
                /db_xref="GI:2792310"
                /translation="MAPTGGVGESRYVKTLRQLEVALIGAGLIARLHLEAWLGAGAAV
                RVYSDDGRSRELAAEFGAKAAGSLEEALDGADAVDICTPTASHHEIALTAIAAGVGVV
                CEKPLAASAEEAEEIVTAAERAGVRLYAAHDVRFAAPYARLHELVASGRLGEGALGRF
                SFSAYHPRPWTGHASARSGGILTDQLLHGADLAHWVFGDVVRVHACYQGDIATPAPEG
                AVATGTAVLTHASGAISQVVSRWTATPRPPVRVAFHVSGTGGSVSYDSEWPQEVRVVD
                GGAGNFAYGGPSVFDTEMREFATAFAGGPEPRIGAKDALAAVRIIHAAAESAWTGRAV
                ELPVRGAA"
     CDS        1448. .2215
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAC01707.1"
                /db_xref="GI:2792311"
                /translation="MKVAILSSTPQAYAGALRGLPDVEVVAAASWDAFEPVRQAAEAG
                ARVLCEYPPAAKETDLKAMIDAAGDRLTFASPACHGEAFAVVRKGIADGGIGELTTVL
                GSVATSVDGVLGAAAPYLLDLADAVLGGEPAQQVYAQTNIVLSGRIGESAAVLTVRYR
                SGQVASFDCRRHGSATGLPAVTFIGDQGSVQYDAGPQLLGGERPELGGEDLEALMLKD
                FLGAGDGPGPDGQAALRTFRIIQAAYESAHTGQPVDL"
     CDS        2353. .2619
                /codon_start=1
                /transl_table=11
                /product="unknown"
                /protein_id="AAC01708.1"
                /db_xref="GI:2792312"
                /translation="MQDRSVLGDPPKQVEAAMSVTQVKIALSGGPAELAPAHRSVDSG
                ALENTLKIRHGAGYEHFVHGGEFQAVDGCDIAVFRWSHRTKIAE"
     CDS        2652. .3842
                /note="ORF0; similar to Streptomyces Sp. ChoP"
                /codon_start=1
                /transl_table=11
                /product="cytochrome P450 monooxygenase"
                /protein_id="AAC01709.1"
                /db_xref="GI:2792313"
                /translation="MTDAISFEVPWDRTDKFDPPAVFDSLREERPLAKMVYPDGHVGW
                IVSSYELVREVLSDLRFSHSCEVGHFPVTHQGQVIPTHPLIPGMFIHMDPPEHTRYRK
                LLTGEFTVRRASRLIPRAEAVAAEQIEVMRAKGAPADVVMDFAKPLVLRMLGELVGLP
                YEERDRYVPAVTLLHDAEADPAEAAAAYEVAGKFFDEVIERRRQRPQDDLISSLVTED
                LTQEELRNIVTLLLFAGYETTEGALATGVFALLHHTDQLAALRAEPEKLDAAIEELLR
                YLTVNQYHTYRTALEDVKLEGELIKKGDTVTVSLPAANRDPAKFGCPAELDIERDTSG
                HVAFGFGIHQCLGQNLARIELRAGFTALLRAFPELRLAVPADEVPLRLKGSVFSVKKL
                PVSW"
     gene       3883. .18090
                /gene="rifA"
     CDS        3883. .18090
                /gene="rifA"
                /note="RifA"
                /codon_start=1
                /transl_table=11
                /product="polyketide synthase"
                /protein_id="AAC01710.1"
                /db_xref="GI:2792314"
                /translation="MRTDLIKPLHVALLENATRFAGKPAFADDHRTVTYGDLEARTRR
                LAGHLAGLGVRHGDRVAICLGNRVSTVESYFAILRAGAVGVPLNPGSATAELEHPLTD
                SGATVVVTDAAQAARLRLAPHVELLVTGDDVPEGAHSYDELALSEPAEPAADDLELDE
                PAWMFYTSGTTGRPKGVVSTQRNCLWSVASCYVPFPGLSDQDRVLWPLPLFHSLSHIA
                CVLSATVVGASVRIADGSSADDVMRLIEAESSTFLAGVPTTYHHLVRAARQRGFSAPS
                LRIGLAGGAVLGAGLRSEFEETFGVPLIDAYGSTETCGAITMNPPDGARVEGSCGLAV
                PGVDVRVVDPDTGLDVPAGEEGEVWVSGPNVMLGYHNSPEATAAAMRDGWFRTGDLAR
                RDDAGYFTICGRIKELIIRGGANIHPGEVEAVLRTVDGVADAAVGGVPHDTLGEVPVA
                YVIPGPTGFDPAALIEKCREQLSAYKVPDRILEVAHIPRTASGKIRRGLLTDEPAQLR
                YAATEHEEQSRHADESVAAALRARLSGLDERAQCELLEDLVRTQAADVLGQPVPDGRA
                FRDLGFTSLAIVELRNRLTEHTGLWLPASAVFDHPTPAALAARVRAELLGITQAVAEP
                VVAADPGEPIAIVGMACRLPGGVASPEDLWRLVAERVDAVSEFPGDRGWDLDSLIDPD
                RERAGTSYVGQGGFLHDAGEFDAGFFGISPREAVAMDPQQRLLLETSWEALENAGVDP
                IALKGTDTGVFSGLMGQGYGSGAVAPELEGFVTTGVASSVASGRVSYVLGLEGPAVTV
                DTACSSSLVAMHLAAQALRQGECSMALAGGVTVMATPGSFVEFSRQRALAPDGRCKAF
                AAAADGTGWSEGVGVVVLERLSVARERGHRILAVLRGSAVNQDGASNGLTAPNGLSQQ
                RVIRRALAAAGLAPSDVDVVEAHGTGTTLGDPIEAQALLATYGQERKQPLWLGSLKSN
                IGHAQAAAGVAGVIKMVQALRHETLPPTLHVDKPTLEVDWSAGAIELLTEARAWPRNG
                RPRRAGVSSFGVSGTNAHLILEEAPAEEPVAAPELPVVPLVVSARSTESLSGQAERLA
                SLLEGDVSLTEVAGALVSRRAVLDERAVVVAGSREEAVTGLRALNTAGSGTPGKVVWV
                FPGQGTQWAGMGRELLAESPVFAERIAECAAALAPWIDWSLVDVLRGEGDLGRVDVLQ
                PACFAVMVGLAAVWESVGVRPDAVVGHSQGEIAAACVSGALSLEDAAKVVALRSQAIA
                AELSGRGGMASVALGEDDVVSRLVDGVEVAAVNGPSSVVIAGDAHALDATLEILSGEG
                IRVRRVAVDYASHTRHVEDIRDTLAETLAGISAQAPAVPFYSTVTSEWVRDAGVLDGG
                YWYRNLRNQVRFGAAATALLEQGHTVFVEVSAHPVTVQPLSELTGDAIGTLRREDGGL
                RRLLASMGELFVRGIDVDWTAMVPAAGWVDLPTYAFEHRHYWLEPAEPASAGDPLLGT
                VVSTPGSDRLTAVAQWSRRAQPWAVDGLVPNAALVEAAIRLGDLAGTPVVGELVVDAP
                VVLPRRGSREVQLIVGEPGEQRRRPIEVFSREADEPWTRHAHGTLAPAAAAVPEPAAA
                GDATDVTVAGLRDADRYGIHPALLDAAVRTVVGDDLLPSVWTGVSLLASGATAVTVTP
                TATGLRLTDPAGQPVLTVESVRGTPFVAEQGTTDALFRVDWPEIPLPTAETADFLPYE
                ATSAEATLSALQAWLADPAETRLAVVTGDCTEPGAAAIWGLVRSAQSEHPGRIVLADL
                DDPAVLPAVVASGEPQVRVRNGVASVPRLTRVTPRQDARPLDPEGTVLITGGTGTLGA
                LTARHLVTAHGVRHLVLVSRRGEAPELQEELTALGASVAIAACDVADRAQLEAVLRAI
                PAEHPLTAVIHTAGVLDDGVVTELTPDRLATVRRPKVDAARLLDELTREADLAAPVLF
                SSAAGVLGNPGQAGYAAANAELDALARQRNSLDLPAVSIAWGYWATVSGMTEHLGDAD
                LRRNQRIGMSGLPADEGMALLDAAIATGGTLVAAKFDVAALRATAKAGGPVPPLLRGL
                APLPRRAAAKTASLTERLAGLAETEQAAALLDLVRRHAAEVLGHSGAESVHSGRTFKD
                AGFDSLTAVELRNRLAAATGLTLSPAMIFDYPKPPALADHLRAKLFGSAANRPAEIGT
                AAAEEPIAIVAMACRFPGGVHSPEDLWRLVADGADAVTEFPADRGWDTDRLYHEDPDH
                EGTTYVRHGAFLDDAAGFDAAFFGISPNEALAMDPQQRLLLETSWELFERAAIDPTTL
                AGQDIGVFAGVNSHDYSMRMHRAAGVEGFRLTGGSASVLSGRVAYHFGVEGPAVTVDT
                ACSSSLVALHMAVQALQRGECSMALAGGVMVMGTVETFVEFSRQRGLAPDGRCKAFAD
                GADGTGWSEGVGLLLVERLSEAQRRGHQVLAVVRGSAVNSDGASNGLTAPNGPSQQRV
                IRKALAAAGLSTSDVDAVEAHGTGTTLGDPIEAEALLATYGQNRETPLWLGSVKSNLG
                HTQAAAGVAGVIKMVMAMRHGVLPRTLHVDRPSSYVDWSAGAVELLTEARDWVSNGHP
                RRAGVSSFGIGGTNAHVVLEEVAAPITTPQPEPAEFLVPVLVSARTAAGLRGQAGRLA
                AFLGDRTDVRVPDAAYALATTRAQLDHRAVVLASDRAQLCADLAAFGSGVVTGTPVDG
                KLAVLFTGQGSQWAGMGRELAETFPVFRDAFEAACEAVDTHLRERPLREVVFDDSALL
                DQTMYTQGALFAVETALFRLFESWGVRPGLLAGHSIGELAAAHVSGVLDLADAGELVA
                ARGRLMQALPAGGAMVAVQATEDEVAPLLDGTVCVAAVNGPDSVVLSGTEAAVLAVAD
                ELAGRGRKTRRLAVSHAFHSPLMEPMLDDFRAVAERLTYRAGSLPVVSTLTGELAALD
                SPDYWVGQVRNAVRFSDAVTALGAQGASTFLELGPGGALAAMALGTLGGPEQSCVATL
                RKNGAEVPDVLTALAELHVRGVGVDWTTVLDEPATAVGTVLPTYAFQHQRFWVDVDET
                AAVSVTPPPAEPIVDRPVQDVLELVRESAAVVLGHRDAGSFDLDRSFKDHGFDSLSAV
                KLRNRLRDFTGVELPSTLIFDYPNPAVLADHLRAELLGERPAAPAPVTRDVSDEPIAI
                VGMSTRLPGGADSPEELWKLVAEGRDAVSGFPVDRGWDLDGLYHPDPAHAGTSYTRSG
                GFLHDAAQFDAGLFGISPREALAMDPQQRLLLETSWEALERAGVDPLSARGSDVGVFT
                GIVHHDYVTRLREVPEDVQGYTMTGTASSVASGRVAYVFGFEGPAVTVDTACSSSLVA
                MHLAAQALRQGECSMALAGGATVMASPDAFLEFSRQRGLSADGRCKAYAEGADGTGWA
                EGVGVVVLERLSVARERGHRVLAVLRGSAVNQDGASNGLTAPNGPSQQRVIRGALASA
                GLAPSDVDVVEGHGTGTALGDPIEVQALLATYGQEREQPLWLGSLKSNLGHTQAAAGV
                VGVIKMIMAMRHGVMPATLHVDERTSQVDWSAGAIEVLTEAREWPRTGRPRRAGVSSF
                GASGTNAHLIIEEGPAEEAVDEEVASVVPLVVSARSAGSLAGQAGRLAAVLENESLAG
                VAGALVSGRATLNERAVVIAGSRDEAQDGLQALARGENAPGVVTGTAGKPGKVVWVFP
                GQGSQWMGMGRDLLDSSPVFAARIKECAAALEQWTDWSLLDVLRGDADLLDRVDVVQP
                ASFAMMVGLAAVWTSLGVTPDAVLGHSQGEIAAACVSGALSLDDAAKVVALRSQAIAG
                ELAGRGGMASVALSEEDAVARLTPWANRVEVAAVNSPSSVVIAGDAQALDEALEALAG
                DGVRVRRVAVDYASHTRHVEAIAETLAKTLAGIDARVPAIPFYSTVLGTWIEQAVVDA
                GYWYRNLRQQVRFGPSVADLAGLGHTVFVEISAHPVLVQPLSEISDDAVVTGSLRRDD
                GGLRRLLASAAELYVRGVAVDWTAAVPAAGWVDLPTYAFDRRHFWLHEAETAEAAEGM
                DGEFWTAIEQSDVDSLAELLELVPEQRGALSTVVPVLAQWRDRRRERSTAEKLRYQVT
                WQPLEREAAGVPGGRWLAVVPAGTTDALLKELTGQGLDIVRLEIEEASRAQLAEQLRN
                VLAEHDLTGVLSLLALDGGPADAAEITASTLALVQALGDTTTSAPLWCLTSGAVNIGI
                QDAVTAPAQAAVWGLGRAVALERLDRWGGLVDLPAAIDARTAQALLGVLNGAAGEDQL
                AVRRSGVYRRRLVRKPVPESATSRWEPRGTVLVTGGAEGLGRHASVWLAQSGAERLIV
                TGTDGVDELTAELAEPGTTVEFCADTDRDAIAQLVADSEVTAVVHAADIAQTSSVDDT
                GVADLDEVFAAKVTTAVWLDQLFEDTPLDAFVVFSSIAGIWGGGGQGPAGAANAVLDA
                LVEWRRARGLKATSIAWGALDQIGIGMDEAALAQLRRRGVIPMAPPLAVTAMVQAVAG
                NEKAVAVADMDWAAFIPAFTSVRPSPLPADLPEAKAILRAAQDDGEDGDTASSLADSL
                RAVPDAEQNRILLKLVRGHASTVLGHSGAEGIGPRQAFQEVGFDSLAAVNLRNSLHAA
                TGLRLPATLIFDYPTPEALVGYLRVELLREADDGLDGREDDLRRVLAAVPFARFKEAG
                VLDTLLGLADTGTEPGTDAETTEAAPAADDAELIDALDISGLVQRALGQTS"
misc_feature    3883. .5775
                /gene="rifA"
                /note="encodes polyketide synthase loading domain"
```

```
                     DQLTLLLDLVRAQVAAVLGHADASAVRVDTAFKDAGFDSLTAVELRNRMRTATGLKLP
                     ATLVFDYPNPQALARHLRDELGGAAQTPVTTAAAKADLDEPIAIVGMACRLPGGVAGP
                     EDLWRLVAEGRDAVSSFPTDRGWDTDSLYDPDPARPGKTYTRHGGFLHEAGLFDAGFF
                     GISPREAVAMDPQQRLLLEASWEAMEDAGVDPLSLKGNDVGVFTGMFGQGYVAPGDSV
                     VTPELEGFAGTGGSSSVASGRVSYVFGFEGPAVTIDSACSSSLVAMHLAAQSLRQGEC
                     SMALAGGATVMANPGAFVEFSRQRGLAVDGRCKAFAAAADGTGWAEGVGVVILERLSV
                     ARERGHRILAVLRGSAVNQDGASNGLTAPNGPSQQRVIRRALVSAGLAPSDVDVVEAH
                     GTGTTLGDPIEAQALLATYGKDRESPLWLGSLKSNIGHAQAAAGVAGVIKMVQALRHE
                     VLPPTLHVDRPTPEVDWSAGAVELLTEAREWPRNGRPRRAGVSAFGVSGTNAHLILEE
                     APAEEPVPTPEVPLVPVVVSARSRASLAGQAGRLAGFVAGDASLAGVARALVTNRAAL
                     TERAVMVVGSREEAVTNLEALARGEDPAAVVTGRAGSPGKLVWVFPGQGSQWIGMGRE
                     LLDSSPVFAERVAECAAALEPWIDWSLLDVLRGESDLLDRVDVVQPASFAMMVGLAAV
                     WQSVGVRPDAVVGHSQGEIAAACVSGALSLQDAAKVVALRSQAIATRLAGRGGMASVA
                     LSEEDATAWLAPWADRVQVAAVNSPASVVIAGEAQALDEVVDALSGQEVRVRRVAVDY
                     GSHTNQVEAIEDLLAETLAGIEAQAPKVPFYSTLIGDWIRDAGIVDGGYWYRNLRNQV
                     GFGPAVAELVRQGHGVFVEVSAHPVLVQPLSELSDDAVVTGSLRREDGGLRRLLTSMA

  Query Match             35.0%;  Score 90.2;  DB 1;  Length 53784;
  Best Local Similarity   63.0%;  Pred. No. 8.4e-06;
  Matches  162;  Conservative    0;  Mismatches   83;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||| ||| |||  ||||  || | ||||   ||||||||||||||||||| || || ||
Db      25231 CTGGCCGGGCTCGCCCCGGCGGACCAGCTCACGCTGCTCCTGGACCTGGTCCGGGCGCAG 25290

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
                |  ||| |||||     ||| ||               ||||   |  | | ||
Db      25291 GTCGCGGCCGTGCTCGGGCACGCCGA-----------CGCGAGCGCCGTCCGCGTCGAC 25338

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              .  ||   |||    ||  |||||||||||  |. |  |  ||||| |||||||||||| |
Db      25339 ACGGCCTTCAAGGACGCCGGCTTCGACTCGCTGACCGCGGTCGAGCTGCGCAACCGCATG 25398

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              | |  ||||| |||||  | ||||||| ||||||| ||||||| || | |  ||
Db      25399 CGGACCGCCACCGGCCTGAAGCTGCCCGCGACGCTCGTCTTCGACTACCCGAACCCCCAG 25458

Qy        241 GCGGTCGCCGCGCGCCT 257
              ||| |||||  || |||
Db      25459 GCGCTCGCCCGGCACCT 25475


RESULT 11
A69720
LOCUS       A69720                 53789 bp    DNA     linear   PAT 07-MAY-1999
DEFINITION  Sequence 3 from Patent WO9807868.
ACCESSION   A69720
VERSION     A69720.1  GI:4774328
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 53789)
  AUTHORS   Schupp,T., Toupet,C. and Engel,N.
  TITLE     RIFAMYCIN BIOSYNTHESIS GENE CLUSTER
  JOURNAL   Patent: WO 9807868-A 3 26-FEB-1998;
            CIBA GEIGY AG (CH)
FEATURES             Location/Qualifiers
     source          1. .53789
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE COUNT     6707 a  19183 c  20504 g   7395 t
ORIGIN

  Query Match             35.0%;  Score 90.2;  DB 6;  Length 53789;
  Best Local Similarity   63.0%;  Pred. No. 8.4e-06;
  Matches  162;  Conservative    0;  Mismatches   83;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||| ||| |||  ||||  || | ||||   ||||||||||||||||||| || || ||
Db      25231 CTGGCCGGGCTCGCCCCGGCGGACCAGCTCACGCTGCTCCTGGACCTGGTCCGGGCGCAG 25290

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
                |  ||| |||||     ||| ||               ||||   |  | | ||
Db      25291 GTCGCGGCCGTGCTCGGGCACGCCGA-----------CGCGAGCGCCGTCCGCGTCGAC 25338

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              .  ||   |||    ||  |||||||||||  |. |  |  ||||| |||||||||||| |
Db      25339 ACGGCCTTCAAGGACGCCGGCTTCGACTCGCTGACCGCGGTCGAGCTGCGCAACCGCATG 25398

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              | |  ||||| |||||  | ||||||| ||||||| ||||||| || | |  ||
Db      25399 CGGACCGCCACCGGCCTGAAGCTGCCCGCGACGCTCGTCTTCGACTACCCGAACCCCCAG 25458

Qy        241 GCGGTCGCCGCGCGCCT 257
              ||| |||||  || |||
Db      25459 GCGCTCGCCCGGCACCT 25475


RESULT 12
AF040570
LOCUS       AF040570               90445 bp    DNA     linear   BCT 16-APR-2001
DEFINITION  Amycolatopsis mediterranei rifamycin biosynthetic gene cluster.
ACCESSION   AF040570 AF040571
VERSION     AF040570.2  GI:12331604
KEYWORDS    .
SOURCE      Amycolatopsis mediterranei.
  ORGANISM  Amycolatopsis mediterranei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
            Amycolatopsis.
REFERENCE   1  (bases 1 to 90445)
  AUTHORS   August,P.R., Tang,L., Yoon,Y.J., Ning,S., Mueller,R., Yu,T.W.,
            Taylor,M., Hoffmann,D., Kim,C.G., Zhang,X., Hutchinson,C.R. and
            Floss,H.G.
  TITLE     Biosynthesis of the ansamycin antibiotic rifamycin: deductions from
            the molecular analysis of the rif biosynthetic gene cluster of
            Amycolatopsis mediterranei S699
  JOURNAL   Chem. Biol. 5 (2), 69-79 (1998)
  MEDLINE   98174059
   PUBMED   9512878
REFERENCE   2  (bases 1 to 90445)
  AUTHORS   Kim,C.G., Yu,T.W., Fryhle,C.B., Handa,S. and Floss,H.G.
  TITLE     3-Amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the
            formation of the precursor of mC7N units in rifamycin and related
            antibiotics
  JOURNAL   J. Biol. Chem. 273 (11), 6030-6040 (1998)
  MEDLINE   98165773
   PUBMED   9497318
REFERENCE   3  (bases 1 to 90445)
  AUTHORS   Yu,T.W., Muller,R., Muller,M., Zhang,X., Draeger,G., Kim,C.G.,
            Leistner,E. and Floss,H.G.
  TITLE     Mutational analysis and reconstituted expression of the
            biosynthetic genes involved in the formation of
            3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin
            biosynthesis in amycolatopsis Mediterranei S699
  JOURNAL   J. Biol. Chem. 276 (16), 12546-12555 (2001)
  MEDLINE   21201076
   PUBMED   11278540
REFERENCE   4  (bases 1 to 90445)
  AUTHORS   August,P.R., Tang,L., Yoon,Y.J., Ning,S., Mueller,R.,
            Hutchinson,C.R. and Floss,H.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1997) Dept. of Chemistry, University of
            Washington, Box 351700, Seattle, WA 98195-1700, USA
REFERENCE   5  (bases 1 to 90445)
  AUTHORS   August,P.R., Tang,L., Yoon,Y.J., Ning,S., Mueller,R.,
            Hutchinson,C.R. and Floss,H.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JAN-2001) Dept. of Chemistry, University of
            Washington, Box 351700, Seattle, WA 98195-1700, USA
  REMARK    Sequence update by submitter
COMMENT     On or before Jan 22, 2001 this sequence version replaced
            gi:2792340, gi:2832263.
FEATURES             Location/Qualifiers
     source          1. .90445
```

SGATVVVTDAAQAARLRLAPHVELLVTGDDVPEGAHSYDELALSEPAEPAADDLELDE
PAWMFYTSGTTGRPKGVVSTQRNCLWSVASCYVPFPGLSDQDRVLWPLPLFHSLSHIA
CVLSATVVGASVRIADGSSADDVMRLIEAESSTFLAGVPTTYHHLVRAARQRGFSAPS
LRIGLAGGAVLGAGLRSEFEETFGVPLIDAYGSTETCGAITMNPPDGARVEGSCGLAV
PGVDVRVVDPDTGLDVPAGEEGEVWVSGPNVMLGYHNSPEATAAAMRDGWFRTGDLAR
RDDAGYFTICGRIKELIIRGGANIHPGEVEAVLRTVDGVADAAVGGVPHDTLGEVPVA
YVIPGPTGFDPAAALIEKCREQLSAYKVPDRILEVAHIPRTASGKIRRGLLTDEPAQLR
YAATEHEEQSRHADESVAAALRARLSGLDERAQCELLEDLVRTQAADVLGQPVPDGRA
FRDLGFTSLAIVELRNRLTEHTGLWLPASAVFDHPTPAALAARVRAELLGITQAVAEP
VVAADPGEPIAIVGMACRLPGGVASPEDLWRLVAERVDAVSEFPGDRGWDLDSLIDPD
RERAGTSYVGQGGFLHDAGEFDAGFFGISPREAVAMDPQQRLLLETSWEALENAGVDP
IALKGTDTGVFSGLMGQGYGSGAVAPELEGFVTTGVASSVASGRVSYVLGLEGPAVTV
DTACSSSLVAMHLAAQALRQGECSMALAGGVTVMATPGSFVEFSRQRALAPDGRCKAF
AAAADGTGWSEGVGVVVLERLSVARERGHRILAVLRGSAVNQDGASNGLTAPNGLSQQ
RVIRRALAAAGLAPSDVDVVEAHGTGTTLGDPIEAQALLATYGQERKQPLWLGSLKSN
IGHAQAAAGVAGVIKMVQALRHETLPPTLHVDKPTLEVDWSAGAIELLTEARAWPRNG
RPRRAGVSSFGVSGTNAHLILEEAPAEEPVAAPELPVVPLVVSARSTESLSGQAERLA
SLLEGDVSLTEVAGALVSRRAVLDERAVVVAGSREEAVTGLRALNTAGSGTPGKVVWV
FPGQGTQWAGMGRELLAESPVFAERIAECAAALAPWIDWSLVDVLRGEGDLGRVDVLQ
PACFAVMVGLAAVWESVGVRPDAVVGHSQGEIAAACVSGALSLEDAAKVVALRSQAIA
AELSGRGGMASVALGEDDVVSRLVDGVEVAAVNGPSSVVIAGDAHALDATLEILSGEG
IRVRRVAVDYASHTRHVEDIRDTLAETLAGISAQAPAVPFYSTVTSEWVRDAGVLDGG
YWYRNLRNQVRFGAAATALLEQGHTVFVEVSAHPVTVQPLSELTGDAIGTLRREDGGL
RRLLASMGELFVRGIDVDWTAMVPAAGWVDLPTYAFEHRHYWLEPAEPASAGDPLLGT
VVSTPGSDRLTAVAQWSRRAQPWAVDGLVPNAALVEAAIRLGDLAGTPVVGELVVDAP
VVLPRRGSREVQLIVGEPGEQRRRPIEVFSREADEPWTRHAHGTLAPAAAAVPEPAAA
GDATDVTVAGLRDADRYGIHPALLDAAVRTVVGDDLLPSVWTGVSLLASGATAVTVTP
TATGLRLTDPAGQPVLTVESVRGTPFVAEQGTTDALFRVDWPEIPLPTAETADFLPYE
ATSAEATLSALQAWLADPAETRLAVVTGDCTEPGAAAIWGLVRSAQSEHPGRIVLADL
DDPAVLPAVVASGEPQVRVRNGVASVPRLTRVTPRQDARPLDPEGTVLITGGTGTLGA
LTARHLVTAHGVRHLVLVSRRGEAPELQEELTALGASVAIAACDVADRAQLEAVLRAI
PAEHPLTAVIHTAGVLDDGVVTELTPDRLATVRRPKVDAARLLDELTREADLAAFVLF
SSAAGVLGNPGQAGYAAANAELDALARQRNSLDLPAVSIAWGYWATVSGMTEHLGDAD
LRRNQRIGMSGLPADEGMALLDAAIATGGTLVAAKFDVAALRATAKAGGPVPPLLRGL
APLPRRAAAKTASLTERLAGLAETEQAAALLDLVRRHAAEVLGHSGAESVHSGRTFKD
AGFDSLTAVELRNRLAAATGLTLSPAMIFDYPKPPALADHLRAKLFGSAANRPAEIGT
AAAEEPIAIVAMACRFPGGVHSPEDLWRLVADGADAVTEFPADRGWDTDRLYHEDPDH
EGTTYVRHGAFLDDAAGFDAAFFGISPNEALAMDPQQRLLLETSWELFERAAIDPTTL
AGQDIGVFAGVNSHDYSMRMHRAAGVEGFRLTGGSASVLSGRVAYHFGVEGPAVTVDT
ACSSSLVALHMAVQALQRGECSMALAGGVMVMGTVETFVEFSRQRGLAPDGRCKAFAD
GADGTGWSEGVGLLLVERLSEAQRRRGHQVLAVVRGSAVNSDGASNGLTAPNGPSQQRV
IRKALAAAGLSTSDVDAVEAHGTGTTLGDPIEAEALLATYGQNRETPLWLGSVKSNLG
HTQAAAGVAGVIKMVMAMRHGVLPRTLHVDRPSSYVDWSAGAVELLTEARDWVSNGHP
RRAGVSSFGIGGTNAHVVLEEVAAPITTPQPEPAEFLVPVLVSARTAAGLRGQAGRLA
AFLGDRTDVRVPDAAYALATTRAQLDHRAVVLASDRAQLCADLAAFGSGVVTGTPVDG
KLAVLFTGQGSQWAGMGRELAETFPVFRDAFEAACEAVDTHLRERPLREVVFDDSALL
DQTMYTQGALFAVETALFRLFESWGVRPGLLAGHSIGELAAAHVSGVLDLADAGELVA
ARGRLMQALPAGGAMVAVQATEDEVAPLLDGTVCVAAVNGPDSVVLSGTEAAVLAVAD
ELAGRGRKTRRLAVSHAFHSPLMEPMLDDFRAVAERLTYRAGSLPVVSTLTGELAALD
SPDYWVGQVRNAVRFSDAVTALGAQGASTFLELGPGGALAAMALGTLGGPEQSCVATL
RKNGAEVPDVLTALAELHVRGVGVDWTTVLDEPATAVGTVLPTYAFQHQRFWVDVDET
AAVSVTPPPAEPIVDRPVQDVLELVRESAAVVLGHRDAGSFDLDRSFKDHGFDSLSAV
KLRNRLRDFTGVELPSTLIFDYPNPAVLADHLRAELLGERPAAPAPVTRDVSDEPIAI
VGMSTRLPGGADSPEELWKLVAEGRDAVSGFPVDRGWDLDGLYHPDPAHAGTSYTRSG
GFLHDAAQFDAGLFGISPREALAMDPQQRLLLETSWEALERAGVDPLSARGSDVGVFT
GIVHHDYVTRLREVPEDVQGYTMTGTASSVASGRVAYVFGFEGPAVTVDTACSSSLVA
MHLAAQALRQGECSMALAGGATVMASPDAFLEFSRQRGLSADGRCKAYAEGADGTGWA
EGVGVVVLERLSVARERGHRVLAVLRGSAVNQDGASNGLTAPNGPSQQRVIRGALASA
GLAPSDVDVVEGHGTGTALGDPIEVQALLATYGQEREQPLWLGSLKSNLGHTQAAAGV
VGVIKMIMAMRHGVMPATLHVDERTSQVDWSAGAIEVLTEAREWPRTGRPRRAGVSSF
GASGTNAHLIIEEGPAEEAVDEEVASVVPLVVSARSAGSLAGQAGRLAAVLENESLAG
VAGALVSGRATLNERAVVIAGSRDEAQDGLQALARGENAPGVVTGTAGKPGKVVWVFP
GQGSQWMGMGRDLLDSSPVFAARIKECAAALEQWTDWSLLDVLRGDADLLDRVDVVQP
ASFAMMVGLAAVWTSLGVTPDAVLGHSQGEIAAACVSGALSLDDAAKVVALRSQAIAG
ELAGRGGMASVALSEEDAVARLTPWANRVEVAAVNSPSSVVIAGDAQALDEALEALAG
DGVRVRRVAVDYASHTRHVEAIAETLAKTLAGIDARVPAIPFYSTVLGTWIEQAVVDA
GYWYRNLRQQVRFGPSVADLAGLGHTVFVEISAHPVLVQPLSEISDDAVVTGSLRRDD
GGLRRLLASAAELYVRGVAVDWTAAVPAAGWVDLPTYAFDRRHFWLHEAETAEAAEGM
DGEFWTAIEQSDVDSLAELLELVPEQRGALSTVVPVLAQWRDRRRERSTAEKLRYQVT
WQPLEREAAGVPGGRWLAVVPAGTTDALLKELTGQGLDIVRLEIEEASRAQLAEQLRN
VLAEHDLTGVLSLLALDGGPADAAEITASTLALVQALGDTTTSAPLWCLTSGAVNIGI
QDAVTAPAQAAVWGLGRAVALERLDRWGGLVDLPAAIDARTAQALLGVLNGAAGEDQL
AVRRSGVYRRRLVRKPVPESATSRWEPRGTVLVTGGAEGLGRHASVWLAQSGAERLIV

TGTDGVDELTAELAEFGTTVEFCADTDRDAIAQLVADSEVTAVVHAADIAQTSSVDDT
GVADLDEVFAAKVTTAVWLDQLFEDTPLDFAVFVFSSIAGIWGGGGQGPAGAANAVLDA
LVEWRRARGLKATSIAWGALDQIGIGMDEAALAQLRRRGVIPMAPPLAVTAMVQAVAG
NEKAVAVADMDWAAFIPAFTSVRPSPLFADLPEAKAILRAAQDDGEDGDTASSLADSL
RAVPDAEQNRILLKVRGHASTVLGHSGAEGIGPRQAFQEVGFDSLAAVNLRNSLHAA
TGLRLPATLIFDYPTPEALVGYLRVELLREADDGLDGREDDLRRVLAAVPFARFKEAG
VLDTLLGLADTGTEPGTDAETTEAAPAADDAELIDALDISGLVQRALGQTS"
CDS 15550. .30759
/function="b-ketoacyl-ACP-synthase, acyltransferase,
dehydratase, ketoreductase, acylcarrier protein"
/note="ORF2; PKS modules 4-6"
/codon_start=1
/transl_table=11
/product="rifamycin polyketide synthase, type 1"
/protein_id="CAA11036.1"
/db_xref="GI:2764762"
/db_xref="SPTREMBL:O52789"
/translation="MANQSWRKNMSAPNEQIVDALRASLKENVRLQQENSALAAAAAE
PVAIVSMACRYAGGIRGPEDFWRVVSEGADVYTGFPEDRGWDVEGLYHPDPDNPGTTY
VREGAFLQDAAQFDAGFFGISPREALAMDPQQRQLLEVSWETLERAGIDPHSVRGSDI
GVYAGVVHQDYAPDLSGFEGFMSLERALGTAGGVASGRVAYTLGLEGPAVTVDTMCSS
SLVAIHLAAQALRRGECSMALAGGSTVMATPGGFVGFARQRALAFDGRCKSYAAAADG
SGWAEGVGVLLLERLSVARERGHQVLAVIRGSAVNQDGASNGLTAPNGPAQQRVIRKA
LASAGLTPSDVDTVEGHGTGTVLGDPIEVQALLATYGQGRDPQQPLWLGSVKSVVGHT
QAASGVAGVIKMVQSLRHGQLPATQHVDAPTPQVDWSAGAIELLAEGREWPRNGHPRR
GGISSFGASGTNAHMILEEAPEDEPVTEAPAPTGVVPLVVSAATAASLAAQAGRLAEV
GDVSLADVAGTLVSGRAMLSERAVVVAGSHEEAVTGLRALARGESAPGLLSGRGSGVP
GKVVWVFPGQGTQWAGMGRELLDSSEVFAARIAECETALGRWVDWSLTDVLRGEADLL
DRVDVVQPASFAVMVGLAAVWASLGVEPEAVVGHSQGEIAAACVSGALSLEDAAKVVA
LRSQAIAASLAGRGGMASVALSEEDATARLEPWAGRVEVAAVNGPTSVVIAGDAEALD
EALDALDDQGVRIRRVAVDYASHTRHVEAARDALAEMLGGIRAQAPEVPFYSTVTGGW
VEDAGVLDGGYWYRNLRRQVRFGPAVAELIEQGHRVFVEVSAHPVLVQPINELVDDTE
AVVTGTLRRREDGGLRRLLASAAELFVRGVTVDWSGVLPPSRRVELPTYAFDHQHYWLQ
MGGSATDAVSLGLAGADHPLLGAVVPLPQSDGLVFTSRLSLKSHPWLAGHAIGGVVLI
PGTVYVDLALRAGDELGFGVLEELVIEAPLVLGERGGVRVQVAVSGPNETGSRAVDVF
SMREDGDEWTRHATGLLGASTSREPSRFDFAAWPPAGAEPIDVENFYTDLTERGYAYS
GAFQGMRAVWRRGDEVFAEVALPDDHREDAGKFGLHPALLDAALHTNAFANPDDDRSV
LPFAWNGLVLHAVGASALRVRVAPGGPDALTFQAADETGGLVVTMDSLVSREVSAAQL
ETAAGEERDSLFQVDWIEVPATETAATEHAEVLEAFGEAAPLELTSRVLEAVQSWLAD
AADEARLVVVTRGAVREVTDPAGAAVWGLVRAAQAENPGRIILVDTDGDVPLGAVLAS
GEPQLAVRGNAFSVPRLARATGEVPEAPAVFSPEGTVLLTGGTGSLGGLVAKHLVARH
GVRRLVLASRRGVAAEDLVTELTEQGATVSVVACDVSDRDQVAALLAEHRPTGIVHLA
GLLDDGVIGALNRERLAGVFAPKVDAVQHLDELTRDLGLDAFVVFSSAAALMGSAGQG
NYAAANAFLDGLMAGRRAAGLPGVSLAWGLWEQADGLTANLSATDQARMSRGGVLPMT
PAEALDIFDIGLAAEQALLVPIKLDLRTLRGQATAGGEVPHLLRGLVRASRRVTRTAA
ASGGGGLVHKLAGRPAEEQEAVLLGIVQAEAAAVLGFNAPELAQGTRGFSDLGFDSLT
AVELRNRLSAATGVKLPATLVFDYPTPVALARHLREELGETVAGAPATPVTTVADAGE
PIAIVGMACRLPGGVMSPDDLWRMVAEGRDGMSPFPGDRGWDLDGLFDSDPERPGTAY
IRQGGFLHEAALFDPGFFGISPREALAMDPQQRLLLEASWEALERAGIDPTKARGDAV
GVFSGVSIHDYLESLSNMPAELEGFVTTATAGSVASGRVSYTFGFEGPAVTVDTACSS
SLVAIHLAAQALRQGECTMALAGGVAVMGSPIGVIGMSRQRGMAEDGRVKAFADGADG
TVLSEGVGIVVLERLSVARERGHRVLAVLRGSAVNQDGASNGLTAPNGPSQQRVIRSA
LAGAGLQPSEVDVVEAHGTGTALGEPIEAQALLATYGKSRETPLWLGSLKSNIGHTQA
AAGVAAVIKMVQALRQDTLPPTLHVQEPTKQVDWSAGAVELLTEGREWARNGHPRRAG
VSSFGISGTNAHLILEEAPADDTAEADVPDAVVPVVISARSTGSLAGQAGRLAAFLDG
DVPLTRVAGALLSTRATLTDRAVVVAGSAEEARAGLTALARGESASGLVTGTAGMPGK
TVWVFPGQGTQWAGMGRELLEASPVFAERIEECAAALQPWIDWSLLDVLRGEGELDRV
DVLQPACFAVMVGLAAVWASVGVVPDAVLGHSQGEIAAACVSGALSLEDAAKVVALRS
QAIAAELSGRGGMASIQLSHDEVAARLAPWAGRVEIAAVNGPASVVIAGDAEALTEAV
EVLGGRRVAVDYASHTRHVEDIQDTLAETLAGIDAQAPVVPFYSTVAGEWITDAGVVD
GGYWYRNLRNQVGFGPAVAELIEQGHGVFVEVSAHPVLVQPISELTDAVVTGTLRRDD
GGVRRLLTSMAELVFVRGVPVDWATMAPPARVELPTYAFDHQHFWLSPPAVADAPALGL
AGADHPLLGAVLPLPQSDGLVFTSRLSVRTHPWLADGVPAAALVELAVRAGDEAGCPV
LADLTVEKLLVLPESGGLRVQVIVSGERTVEVYSQLEGAEDWIRNATGHLSATAPAHE
AFDFTAWPPAGAQQVDGLWRRGDEIFAEVALPEELDAGAFGIHPFLLDAAVQPVLADD
EQPAEWRSLVLHAAGASALRVRLVPGGALQAADETGGLVLTADSVAGRELSAGKTRAG
SLYRVDWTEVSIADSAVPANIEVVEAFGEEPLELTGRVLEAVQTWLVTAADDARLVVV
TRGAVREVTDPAGAAVWGLVRAAQAENPGRIFLIDTDGEIPALTGDEPEIAVRGGKFF
VPRITRAEPSGAAVFRPDGTVLISGAGALGGLVARRLVERHGVRKLVLASRRGRDADG
VADLVADLAADVSVVACDVSDRAQVAALLDEHRPTAVVHTAGVIDAGVIETLDRDRLA
TVFAPKVDAVRHLDELTRDRDLDAFVVYSSVSAVFMGAGSGSYAAANAFLDGLMANRR
AAGLPGLSLAWGLWDQSTGMAAGTDEATRARMSRRGGLQIMTQAEGMDLFDAALSSAE
SLLVPAKLDLRGVRADAAAGGVVPHMLRGLVRAGRAQARAASTVDNGLAGRLAGLAPA

```
                     /transl_table=11
                     /product="luxR-family transcriptional regulatory protein"
                     /protein_id="BAB69190.1"
                     /db_xref="GI:15823973"
                     /translation="MLASAGQGEGGVVLVNAGPGVGRTALLDAFLRKATESGARVCAA
                     AGSSAETEAEFGVVGQLFPPTGPLGAGFRLARAIGSRDAGPGLDADRLFDALCQELRR
                     RPTVIAVDDAQLADAASLRFLRYLVRRLRTAPVMMVLTEPTGPDALPPSFQAELLRHP
                     QCRRLGLLPLSRDGVTQVVSAQVGSDRVPRLAADFHAISGGNPLLVHALLADHRAAHR
                     LSPVPSVPTVPSVSSASSAASASSAASAASAVSAAPAASAAFAKAALAWAYRDEPVLF
                     DVACGFAVLGESATPARVACLVDRGADAVDQVMTALDTAGLLAGGTFRSALVGKALLE
                     HVDVETRAELRRRAARLLHGDGAPATVVARHLLAAPVAEPWAGRTLLAAAEKAMQEGR
                     EELSLDCLRLAGQSATTEQERAAVVMARVKIGWEIDPRLVSPWLDELVAALRRGDLRS
                     GDAAWIVKHLAWQDRFPEAVESLSTLAAAGDAGEPGVEAELHFVRHWLRHSCPPLEAS
                     VPGPSNPTRPRTPRLRRVSPASYAGEVLGKVLTEGPDDHTVAMAEEILRGCRYSETVV
                     EALEAALLALVYAGRPDRAAFWCDGLLQQAGTRSTGTSVAILSGIRAEIALRQGALPA
                     VEQYARSALSAISLPGWGVAVGSPLAALVHSTTASGMSGVAGAWLNQAVPDGMFRTRH
                     GLLYTHARGHYHLAVDRPAAALDDFLTCGGLAKEWGMDVPTFLPWRTSAAQAQLALGN
                     PGQARALVREQLGRPGGSTARARGVALRLLAATGEPGRRPALLRESVNLLESCGDQVE
                     LLRALADRCQALDHVGAPAKARMTARHARNVADACGGDALFRKLFRDEVPDREGDPDG
                     LTGEEQGSASLTAAERRVTALAALGHSNREIGRKLFITKSTVEQHLTRVYRKLGVRNR
                     ADLGDLLAGTNLGAGTPAMAMASSAVAG"
CDS                  10933. .11271
                     /codon_start=1
                     /transl_table=11
                     /product="P450-like protein"
                     /protein_id="BAB69191.1"
                     /db_xref="GI:15823974"
                     /translation="MEVGGQTIRAGEGVLCTLSTANRDEGAFTDPDELDLRRDARSHL
                     SFGFGIHQCIGQSLARAEPQIALETLLRRLPGLRLAASFEELRFRDGVGFYGVKELPD
                     PCPALPPDRP"
gene                 11458. .29898
                     /gene="olmA1"
CDS                  11458. .29898
                     /gene="olmA1"
                     /codon_start=1
                     /transl_table=11
                     /product="modular polyketide synthase"
                     /protein_id="BAB69192.1"
                     /db_xref="GI:15823975"
                     /translation="MACRLPGAAGPAEFLALLRNGEDAVGAAPGDRRYAPRRGGFLDS
                     VDEFDAGFFGISPREAAAMDPQQRLVLELSWEALEDAGIPPARLGGSGAGVFVGAIAD
                     DYATLSRAAGADAATPQTTTGLNRGIIANRVSYTLGLRGPSFTVDSGQSSSLVAVHLA
                     AESLRRGECSVALAGGVNLILAEDSTVAVERFGALSPDGYCYTFDARANGYVRGEGGG
                     VVVLKPLADAIADGDDIHCILAGSAVNNDGGGEGLTVPDRAGQEAVLTAAYEQAGISP
                     AAVGYVELHGTGTPTGDPVEAAAVGAVLGAGRPAGQPLLVGSVKTNIGHLEGAAGIAG
                     LLKAVLAVRHREIPATLHHASPNPAIPMAELGIGVNTALTPWPEQAGPLVVGVSSFGM
                     GGTNCHVVLTEPAGYDAAARPATAGIRPAGTPAPLLVSGRDEQALRAQARKLRDHLDT
                     HPGLPLRDVAHTLAAARAPLARRAVLLAADREDVSAGLARLADGTPGADLVRGTVSPG
                     ALAFVFTGQGSQRPGMTAELYRTSAEYAAALDEVCASLDPWLRRPLRDVLFAAEGSDE
                     AALLDRTEFTQPALFAVEVALFRYAEHCGLTPRFLLGHSVGELAAAHVAGVLSLADAC
                     TLVAARGRLMQDRPATGAMVSVQATEDELAPFVSESVTLAALNGPASTVIAGDEDAVL
                     AVAAHWAARGRKTKRLRVSHAFHSPHMDGMLTEFRRIAAGLTFGAPRIPIVSNETGTL
                     LTEADARSPEYWVRHARAAVRFLDGVRLLEDQGVTTLLELGPDGVLSAMARDCLRDAG
                     TVCVPLLRGRDRTEPEALLAALAGLYVRGVPMDWERTATEAGARRVPLPTYAFQRRRH
                     WLPEPSASASRPVVNGGGHAGQAGHAGQAGRAGQAGHAGHDGEPAPAVRDRTTPELVR
                     DRTTLELVRDATAQVLGHASPDDVATHEAFRELGFSSLMLAELGERLTEATGRRVPTT
                     LLFDHPTPDALARRLEADGAERTDAGPVSPPAAAAHDDDPIVIVGMGCRLPGGVRSPE
                     DLWRFVAAGGDAISALPVDRGWEFAADFPGEGGFLADVAEFDAGFFGISPREALAMDP
                     QQRLLLETSWEALERAGVDALKLRGSRTGVFVGASPSEYGPRLHETSDADGHVLTGTA
                     PSVLSGRIAYVLGLEGPALTVDTACSSSLVALHLAAQALRGGECDLALAGGVTVMATS
                     GMFAEFARQGGLAPDGRCKAFADGADGTGWGEGVGMLAVQRLSDAVRDGRPVLAVVRG
                     SAVNSDGASNGLTAPNGPSQQRVIRQALASAGLSAADVDMVEAHGTGTSLGDPIEAQA
                     LLATYGQDRPDDRPLWLGSVKSNIGHTQAAAGVAGVIKSVLALRNGLLPKTLHVDEPT
                     PEVDWSAGAVELLTEGREWPETDGPRRAGVSAFGISGTNAHVILEQAPSVEEPVTAPA
                     DDGLLVPWVVSARSEEALRAQARRLADHVRESGLRPADVGLSLAVTRAGLEHRAVLVA
                     EDEAGFLAHLDALAEGRTPAGVVVEGGTAFLFSGQGSQRAGMGRELYSAHPVFAAAFD
                     EVCAAFDGLLPGALREVVFGGAGLDRTEWAQPALFALEVALFELVSSWGVRADVLLGH
                     SIGELAAAYVAGVWSLADACRIVAARGRLMQALPEGGAMVAVEAAEDELPDLPGGVSV
                     AAVNGPRSLVLSGDEEPVTALAQAFAEQGRRTKQLTVSHAFHSARMEPMLAEFAETLA
                     AVEFRPPRIPVVSNVTGEMAGDEFTTPAYWVRHVREAVRFADGVGAAHAAGVRRYLEL
                     GPDGTLTTLAHDTAHDAGRDVAAVPALRGDQPETRTVVTAVGRLYATGLTVDWPAFFA
                     PYGARPTELPTYAFQHRRYWLEPVAPSSAPPANPLRYRTAWASVPDGPRPSLSGTWAV
                     VVPASGSYEDIVAGVTAALRDHGARPVVVEVDPAGPERADVADTLRAALGEDAGGVVS
                     LLGTAGSLSSGVAATAVLVQAWTDLGGGARLWCLTRGAVSVSPSDPLAAPAQAQLWGM
                     GRVAALEHPERWGGLVDLPAVPDDRAWTRLCAVLAGGSGEDQVAVRAAGLFARRLGRA
                     GTRADAAPAGPSPWRTDGTVLVTGGTGALGAHLAQWLAAAGAAHVLLTSRRGPDADGA
                     AELTARLRETGTEVTVAACDVTDRDALAELLSGLPADRPLTGVVHAAGVLDDGVLDSL
                     TPDRFAAVAGPKVTGAWHLHELTRDLDLSAFVLFSSFAGTVGLAGQANYAAANAYLDA
                     LAVHRAQLGLPATAVAWGSWSGAGMAADTEAARRQLARTGLVPLAPDAALAALGQVMD
                     DGETAVTVADVDWDRFAAGFGPGRPSPLLTGVPEVRRIEERETGRAGAPAETDRLGGL
```

Query Match 35.6%; Score 91.8; DB 1; Length 104326;
Best Local Similarity 63.4%; Pred. No. 4e-06;
Matches 163; Conservative 0; Mismatches 82; Indels 12; Gaps 1;

```
Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||||||| |||| | ||  |  ||||||       ||| ||||||||||| ||  |  ||
Db      89627 CTGACCGGGCTGCCTCCAGCCGAGCAGCTTCGCAGGCTGCTGGACCTGGTCCGTTCGAAC 89568

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              |    | |||||   | ||| |      || |   | |||| ||||            |
Db      89567 GCGGCCACCGTGCTCGGCCACGCCTCGGCCGGGGCGATCGCGGCGGA----------G 89520

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              |   |||  | | ||||||||||||||  | |  |  || ||  | |||||||| |||
Db      89519 CAGTCCTTCAGCGAGCTCGGCTTCGACTCGCTGACCGCCGTGGAGTTCCGCAACCGACTC 89460

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              ||  |||||||| |||||| |||| |  |||||| ||||||| ||  | |||||||||
Db      89459 GGCGCGGCCACGGGTCTGCGGCTGCCGGCCACGCTCGTCTTCGAGCATCCGACGCCGGCC 89400

Qy        241 GCGGTCGCCGCGCGCCT 257
              ||| |||||| |   ||
Db      89399 GCGCTCGCCGGGGAACT 89383
```

```
RESULT 10
AMM223012
LOCUS       AMM223012              53784 bp    DNA     linear   BCT 09-FEB-1998
DEFINITION  Amycolatopsis mediterranei genes encoding rifamycin polyketide
            synthases, ORFs 1 to 5.
ACCESSION   AJ223012
VERSION     AJ223012.1  GI:2764760
KEYWORDS    ORF1; ORF2; ORF3; ORF4; ORF5; polyketide synthase; rifamycin.
SOURCE      Amycolatopsis mediterranei.
  ORGANISM  Amycolatopsis mediterranei
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Pseudonocardineae; Pseudonocardiaceae; Amycolatopsis.
REFERENCE   1  (bases 1 to 53784)
  AUTHORS   Schupp,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1997) Schupp T., Core Technology, Novartis Pharma
            AG, CH 4002 Basel, Postfach, Novartis K-681.344, 4002 Basel,
            SWITZERLAND
REFERENCE   2  (bases 1 to 53784)
  AUTHORS   Schupp,T., Toupet,C., Engel,N. and Goff,S.
  TITLE     Cloning and sequence analysis of the putative rifamycin polyketide
            synthase gene cluster from Amycolatopsis mediterranei
  JOURNAL   Unpublished
FEATURES             Location/Qualifiers
     source          1. .53784
                     /organism="Amycolatopsis mediterranei"
                     /strain="LBG A3136"
                     /db_xref="taxon:33910"
     CDS             1336. .15543
                     /function="CoA ligase, acylcarrier protein,
                     b-ketoacyl-ACP-synthase, acyltransferase, ketoreductase"
                     /note="ORF1; PKS modules 1-3"
                     /codon_start=1
                     /transl_table=11
                     /product="rifamycin polyketide synthase, type 1"
                     /protein_id="CAA11035.1"
                     /db_xref="GI:2764761"
                     /db_xref="SPTREMBL:O54666"
                     /translation="MRTDLIKPLHVALLENATRFAGKPAFADDHRTVTYGDLEARTRR
                     LAGHLAGLGVRHGDRVAICLGNRVSTVESYFAILRAGAVGVPLNPGSATAELEHPLTD
```

```
  Best Local Similarity   63.4%;  Pred. No. 4.9e-06;
  Matches  163;  Conservative    0;  Mismatches   82;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||||||| |||  |  |   | |   |||  | ||  ||||||  |  | | ||||||||||
Db      30648 CTGACCGGGCTCACGGCACAGGAAGGGCACGCCCTCGTCCTGGCCGAGATCCGCGCCCAC 30707

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
                |   ||||||||||  | ||| |    ||||    | | |   |||    | |
Db      30708 GCGGCGGCGGTGCTGGGCCACGGCTCCGACGACTCGATCCCCGAGGACCGGGCC------ 30761

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                 |||   ||||||||||||||||  |||  |  || ||  ||||||||| ||
Db      30762 ------TTCAAGGACCTCGGCTTCGACTCGCTCACCGCCGTGGAGATGCGCAACCGGCTG 30815

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              |||  |||||| ||||| ||| ||||||  || ||| |||||.|||| | || |||| |
Db      30816 AGCGCGGCCACCGGCCTCCGGCTGCCCGCCACCCTCGTCTTCGACCACCCGACCCCGGGC 30875

Qy        241 GCGGTCGCCGCGCGCCT 257
              | | | ||||  | |||
Db      30876 GAGCTGGCCGGCCACCT 30892


RESULT 9
AB070940/c
LOCUS       AB070940               104326 bp    DNA     linear   BCT 22-JAN-2002
DEFINITION  Streptomyces avermitilis oligomycin biosynthetic gene cluster.
ACCESSION   AB070940
VERSION     AB070940.1  GI:15823967
KEYWORDS    .
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
  TITLE     Genome sequence of an industrial microorganism Streptomyces
            avermitilis: deducing the ability of producing secondary
            metabolites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
  MEDLINE   21477403
REFERENCE   2  (bases 1 to 104326)
  AUTHORS   Ikeda,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .104326
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            complement(1. .1134)
                     /gene="mcmA"
     CDS             complement(1. .1134)
                     /gene="mcmA"
                     /note="methylmalonyl-CoA mutase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB69185.1"
                     /db_xref="GI:15823968"
                     /translation="MTGYSTAADANARFRQLVGSGTRAVPVVFDLPTRAGLDSDAPAA
                     RGLVGRRGVAVDSIDDMRVLFGGLRLDRVVPSLAVDAPAAPLLVLYQLVAEEHGAPAQ
                     RLAGAVRGDVFKEFLARGPGVFPLRPSLRLAMNVFAYCQAELPRWQCLSVCGHHLAGT
                     GADPALEVAFTVAGGIEYLRAAVTAGLDVRTVAPRLTLCLVGGTTGRENRAKLRATRR
                     VWTRTANERFGSGRTAPRLYVRTPRDGLATLACTEDIEAGATALVDRIERLGGTLTVL
                     EHGLAGQLLGRPHGDRPATPRPHPPHQDSLDLQARQTERLAKLRAWRVQPAVDEALLR
                     LRKAAQGDDNVLYPMREALAARATVGEVCDALRSTWGTGPSAT"
     gene            complement(1131. .1886)
                     /gene="pptA"
     CDS             complement(1131. .1886)
                     /gene="pptA"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphopantheteinyl transferase"
                     /protein_id="BAB69186.1"
                     /db_xref="GI:15823969"
                     /translation="MRAHEVFADVPVSSLHPREAALVQDAVESRRREFATGRHCARRA
                     LASLGVPPGPLLATPRGAPRWPAGVCGAITHCAGYRAAAVARAPGIRAVGIDAEPDRP
                     LPEGVLEAVALPAEQFQVKEHTARHPGVVWGKLLFSAKESVYKAWFPATGRPLGFEEA
                     LIDLDPVSSTFRVRVLPTALRAAEPELRDLHGRWLVRHGIIVTAVVVDGPGRDTPHAP
                     SPYDRTKHSAPTGRSADRPGGRPTDRATDRPTA"
     gene            complement(1910. .2875)
                     /gene="simX4 homolog"
     CDS             complement(1910. .2875)
                     /gene="simX4 homolog"
                     /codon_start=1
                     /transl_table=11
                     /product="SimX4 homolog"
                     /protein_id="BAB69187.1"
                     /db_xref="GI:15823970"
                     /translation="MSQDRHDGSTAALGGPSHGGRLLAVSDIHVRHRQNRDIVQDLHP
                     GCAEDWLIVAGDVAESVADITWALSLLRDRFAKVIWAPGNHELWTAPGDPDELRGERK
                     YRHLVDLCRALGVVTPEDPYPVWHGAGGPVAVAPLFLLYDYSFLPPGAWSKSHALRLA
                     DDAGVICTDEFLLHPDPYPSREAWCRARVRATRARLDALPDDLPTVLVNHFPLVREPT
                     RVLRHPEFALWCGTTATADWHLRYRAAAVVHGHLHIPRTIRCDDVPHHEVSLGYPREW
                     QRYGRRPDWPVQILTPAPAPAPAPVPVPVPAPPGHRTASPADRRP"
     gene            3213. .3971
                     /gene="olmC"
     CDS             3213. .3971
                     /gene="olmC"
                     /codon_start=1
                     /transl_table=11
                     /product="thioesterase"
                     /protein_id="BAB69188.1"
                     /db_xref="GI:15823971"
                     /translation="MSANRVDGNLWVRQFHPAPTAATKLVCLPHAGGSASYFLPVSKA
                     LAPGVDVLAIQYPGRQDRRHEKCIDNIAELADAVTAVLEPLADRPLVLFGHSMGATLA
                     FEVALRLERKGIVPLALFASGRRAPSRHRDESVHLRGDDEIIAELKTLSGTRSEVLGD
                     EEILRMVLPAIRSDYKAAETYRYAGGARLTAPIHAHFGVEDPRVSLDEAQAWADHTTG
                     HFELRSHPGGHFYLNDQAPRVIEEIARVSSSVAS"
     gene            4643. .7438
                     /gene="olmRI"
     CDS             4643. .7438
                     /gene="olmRI"
                     /codon_start=1
                     /transl_table=11
                     /product="luxR-family transcriptional regulatory protein"
                     /protein_id="BAB69189.1"
                     /db_xref="GI:15823972"
                     /translation="MLIGRGQEIDRLRAAYSSVRSTGRSRAVLVESAAGCGKSEILED
                     LVRHVESLGLIALRATAIAEERNTPLSVLRSLVNGAPFPDDLVRRFHTMSDAAARMIS
                     TEGRDKARRDAAWYELMRTFCTEVRALAQRAPLVVAVDDIHHGDADSLRQLLHLARHC
                     RNVPVLLTFTQSLAGGPADRTFGAELLRQPNLEHIRLPRFCRQDIADMLAEQGRADAG
                     QGTVERYFAVSGGNPLLLRALLEDARLEPDDTVDAQDVGPRAGQAYGRAVLACLDRGG
                     EKAHEVAAGLAVLGGAFTHELLSRLLDISAAALTENLRALRAAGIVVGTSFQHPAAKA
                     AVLDSLEPSHRLRLHQRSAALLHASGAPAALAAEHLHEAGQADGAWAVGVLQQAAEQA
                     LLADEARRAVSYLELAHAASADPRGRTEIKIRLGSVTRRVSAADAERHVDDALRGFGA
                     GLLTPSDTGGLAALLLAHGRLEQARSVLEQGRRPGDGDADRTAARRQLAEIRDSVLNR
                     GAWCAGTRAAWHENPAGRPSTGRAGTGLRAWLSGSHAALLAQRIAGTEGLLEITHLTD
                     ATFDLVLNAVWTLLALGAVERALHWCDTFLREAKRRDAPGWEAVFAAVRAEGALRGGN
                     LAEAERDARRCFAVLPERHRSALEGSAVSYLVTVQIARGKYEDAARQLSRPMADSLPD
                     TIHWLDYLRARGLYYMATRQYHSALRDFHEVGRLAERWGTDWPALLPWRTDTAEALLR
                     LGEREQGLRLATEQLAMVQDGAPRIKGISLRLQGLASEPAARLGLLNRAVVELSRADD
                     RLELSRTLFDLADAHRETRQTRQAAMVLRRARQLAKSCGAQPPRDRIGSPDAAGASGE
                     LFGEGPERNADWGHLSESERRVAGLAASGYTNREISVRLYITMSTVEQHLTRVYRKLK
                     INGREELLAEARAGVQV"
     gene            7658. .10486
                     /gene="olmRII"
     CDS             7658. .10486
                     /gene="olmRII"
                     /codon_start=1
```

GTSGSVVSGRVAYTLGLEGPAVTVDTACSSSLVALHLAVQALRGGECDMALAGGVTVM
AGPGMFVEFSRQRGLAADGRCKAFADGADGTAWAEGAGVVLVERLSDARRLGHPVLAV
VCGSAVNQDGASNGLTAPSGPSQERVIRQALGNARLTVADVDVVEAHGTGTRLGDPIE
AQALLGTYGRDRDGGRPVWLGSLKSNIGHAQAAAGVAGVIKMVLAMRYGWLPRTLHVD
EPSRHVDWSAGGVWLLTEAREWPGVDRPRRAAVSAFGVSGTNAHLILEAPDTAEAESA
TTPVRSEVSESAAVLDARSGVVPVVVSGRSRVVVREAAGRLAEVVEAGGVGLADVAVT
MAGRSRFGYRAVVLARGEAELAGRLRALAGGDPDAGVVTGAVVDPETGSGGGGVVLVF
PGQGTQWVGMGAGLLGSSEVFAASMRECARALSVHVEWDLLEVVSGGAGLERVDVVQP
VTWAVMVSLARYWQAMGVDVAAVVGHSQGEIAAATVAGALSLEDAAAVVALRAGLIGR
YLAGRGAMAAVPLPAGEVEAGLAKWPGVEVAAVNGPASTVVSGDRRAVAGYVAVCQAE
GVQARLIPVDYASHSRHVEDLKGELERVLSGIRPRSPRVPVCSTVAGEQPGEPVFDAG
YWFRNLRNRVEFSAVVGGLLEEGHRRFIEVSAHPVLVHAIEQTAEAADRSVHATGTLR
RQDDSPHRLLTSTAEAWAHGATLTWDPALPPGHLTTLPTYPFNHHHYWLDTTPTTPAT
TTQSPTDAWRYRVTWKALTESSPVRPHSIGRCLLVAPPTTDGELLDGLTTVLSERGAS
VARLEVPIGARRAEVAELLKPSMESAGEENTTVVSLLGLVPSTDAVRTSIALLQAVSD
IGVPAARVWALTRRAVAVVPGETPQDAGAQLWGFGRVAALELPDIWGGLIDLPETAEL
TRTPETSQPPQTPERLPQTPNRRALELAAAVLAGRDGEDQVAVRASGIYGRRVSRAAA
AGAASWQPSGTVLITGGMGAIGRRLARRLAAEGAERLVLTSRRGPEAPGAAELAEELR
GHGCEVVHAACDVAERDALAALVTAYPPNAVFHTAGILDDAVIDTLSPESFETVRGAK
VCGAELLHQLTADIKGLDAFVLFSSVTGTWGNAGQGAYAAANAALDALAERRRAAGLP
ATSVAWGLWGGGGMAAGAGEESLSRRGLRAMDPDAAVDALLGAMGRNDVCVTVVDVDW
ERFAPATNAIRPGRLFDTVPEAREALTAAGTTSATPDGAPELARRLSMLNETERLRKL
VELVRTEAAFVLRHPNTDAIGAERPFKSAGFDSLTSLELRNRLNAGTGLKLPATVIFD
HPSPTALARLLLDRLTGAGAPAPAADEPPLPVAVADDDPVVIVGMACRFPGGAGTPEA
LWKLVTEERDVIGAAPTDRGWDLDSVYDPEPGVAGKTYVREGGFLHDAAEFDAEFFGI
SPREAVAMDPQQRLLLETSWEAIERAGIDPHSLHGSRTGVYVGLTHQEYASRLHEAPE
EYEGYLLTGKSASVVSGRISYTLGLEGPSLSIDTACSSSLVALHNAAQALRGGECDMA
LAGGVTVMAAPGLFVEFSRQRGLAADGRCKAFADGADGTAWAEGAGVVLVERLSDARR
LGHPVLAVVCGSAVNQDGASNGLTAPSGPSQERVIRQALANARLTVADVDVVEAHGTG
TRLGDPIEAQALLGTYGRDRDAECPVWLGSLKSNIGHAQAAAGVAGVIKMVLAMRYGW
LPRTLHVDEPSRHVDWSAGGVRLLTEAREWPGVDRPRRAAVSAFGVSGTNAHLILEAP
EALEALEATDAPEAPEAPEAPDVTDVTEALEAPDATEAEGAKAPGSPEEAQPAVGVVP
VVVSGRSRVVVREAAGRLAEVVEAGGVGLADVAVTMAGRSRFGYRAVVLARGEAELAG
RLRALAGGDPDAGVVTGAVVDPETGSGGGGVVLVFPGQGTQWVGMGAGLLGSSEVFAA
SMRECARALSVHVEWDLLEVVSGGAGLERVDVVQPVTWAVMVSLARYWQAMGVDVAAV
VGHSQGEIAAATVAGALSLEDAAAVVALRAGLIGRYLAGRGAMAAVPLPAGEVEAGLA
KWPGVQVAAVNGPASTVVSGDRRAVAGYVAVCQAEGVQARLIPVDYASHSRHVEDLKG
ELERVLSGIRPRSPRVPVCSTVAGEQPGEPVFDAGYWFRNLRNRVEFSAVVGGLLEQG
HRRFIEVSAHPVLVHAIEQTAEAADRSVHATGTLRRQDDSPHRLLTSTAEAWAHGATL
TWDPALPPGHLTTLPTYPFNHHHYWAVTSPAGVGDAAAGRFGMTWEDHPFLRGGLPLA
DSGERVFAGRLAGSEHDWLTDHAVSGVTLLPGTAFVEFALHAGAATGCGRLEELSVEA
PLVLPAAGGVRVQMRVSAADESGRRRVAIHSAPEAAVHSAAEGGDSAGVWTRHGEGTL
VPDPEPTPPDADWARAWPPAGERVEPAELYERFGALGYEYGEAFAGVRAVWRQPDALL
AEVLLPDRASTGAGRFGVHPALLDAALQPWIAGGLLEVPEDAVLLPFAWQGVSLYATG
AGALRVRLTKAGDGAVSLQAADTSGAAVLSLGALVMRPLARRKLDVLLGTDAGERSLY
RVEWQPRLLPAGPPRSWAVLGPDADRLAGTPGLGDQPDGGPTALYPEVRALRKALAAG
APRPEAVVLPVLSGAGATPESVRQTTERCLTALQDWLDAEELVDTPLIVLTRGAVAAV
PGEEIGDLACAGVWGLVRSARSEHPGRFALVDTDGHPDDRTALPLALRAVLDGAGQLS
LRAGTARTPVLLRAGTPEEQRGPAFDPAGTVLVTGATGTLGRLLARHLAAEHGVRHLL
LLSRGGRAAEGADELAAELAGLEAEPCFAACDAADREALARVLAEVPADRPLTGVIHA
AGVLDDGTLDALTPERIGTVMRPKADAALNLHELTRTSPLSVFAVFSGAAGILGRPGQ
ANYAAANTFLDALAQHRRAHGLPAVSLAWGLWGGATGMTGHLSGTDLRRMRRSGIAPM
THDQGLALFDRALAASAEDPLLVPMRLDLAALVRERAEHGPDAVPGPLLGLLPARAAV
RQAAAPVRGGAPAPAGGEGTAERLAGLGEEARLRELVRLVRAEVSGVLGYSGPDAVEP
GRPFKDLGFDSLTAVELRNRLGAATGLRLPTALVFDRPTSQAVAEYLAAELAGPRDGG
DTAAAAFEGLEALAAAVGALAEDDLRRDVLRRRLTELAAALTPQGRNPSAPAPAPSDL
DERLDSANDDDLFAFIEEQL"

misc_feature 942. .3929
/gene="tylG"
/function="encodes starter module"

misc_feature 942. .2156
/gene="tylG"
/function="encodes beta-ketoacyl ACP synthase KS-Q of starter module"

misc_feature 2571. .3557
/gene="tylG"
/function="encodes acyltransferase domain of starter module"

misc_feature 3675. .3929
/gene="tylG"
/function="encodes acyl carrier protein domain of starter module"

misc_feature 3993. .8471
/gene="tylG"
/function="encodes module 1"

misc_feature 3993. .5264
/gene="tylG"
/function="encodes beta-ketoacyl ACP synthase of module 1"

misc_feature 5631. .6617
/gene="tylG"
/function="encodes acyltransferase domain of module 1"

misc_feature 7410. .7949
/gene="tylG"
/function="encodes beta-ketoreductase domain of module 1"

misc_feature 8220. .8471
/gene="tylG"
/function="encodes acyl carrier protein domain of module 1"

misc_feature 8541. .13970
/gene="tylG"
/function="encodes module 2"

misc_feature 8541. .9812
/gene="tylG"
/function="encodes beta-ketoacyl ACP synthase of module 2"

misc_feature 10260. .11246
/gene="tylG"
/function="encodes acyltransferase domain of module 2"

misc_feature 11319. .11876
/gene="tylG"
/function="encodes dehydratase domain of module 2"

misc_feature 12861. .13415
/gene="tylG"
/function="encodes beta-ketoreductase domain of module 2"

misc_feature 13719. .13970
/gene="tylG"
/function="encodes acyl carrier protein domain of module 2"

CDS 14351. .19945
/gene="tylG"
/note="ORF2"
/codon_start=1
/transl_table=11
/product="tylactone synthase module 3"
/protein_id="AAB66505.1"
/db_xref="GI:2317861"
/translation="MTAELVATRKRLGALEERAREPIAVVAMSCRYPGGVTTPEDLWR
LLADERDAVSGLPRDRGWDLDALYDPDGGPGTSYAREGGFLSHCAGFDAEFFGISPRE
ALAMDPQQRLLLETSWEALERAGVTADRARGSRTGVYAGVMYDDYGARVLYGAGAGPP
EDLEGYLVNGSAGSIASGRVSYTFGLRGPAVTVNTACSSSLVSLHLAVRALRNGECDM
ALAGGATVLSTPTVLVDFSRQRGLAPDGRCKAFADSADGTSWAEGAGMLLLQRLSDAR
REGRPVLAVIRGSAVNQDGASNGLTAPNGRAQRQVIEDALRDAGVGPDQVDAVEAHGT
GTELGDPIEAGALLATYGTARTAERPLWLGSLKSNIGHTQAAAGVAGVIKMVLAMRHG
RLPRTLHVDRPTTRVDWEKGGVRLLTEPVPWPGEAGEPRRAGVSSFGASGTNAHVVLE
SVPAGEPPAAGRPEDTGGAWTVSGRGPAALRAQAARLYDALTGTGTGTGQGAGQGAGP
GTAEVAGALAHARTAFRHRAVVLGGNRAELLAGLRELAEEEHPGPRVVTGTAPATERR
TAFLFSGQGSQRAGSGRGLYRRHPVFARALDEVCAALEPHLHRPLRDLMFAEPGSPEA
EPLDRTEFTQPALFALQTALFRLAEHHGLRAEALCGHSVGEIAAAHAAGVLTLPDAAR
LVAARGRLMQALPAGGAMAALRATAEEIAPLLERRAGELALAAVNGPSSVVVSGDEAA
VLELLEQWRAEGREARRLAVSHAFHSPRMDGMLTQFDRVARTLTFAPPTIPLVSTLTG
TPVTEETLCTADHWVRQAREPVRFLDAMRTLRADGIDTFVELGPDGVLSAMARDCADD
RPDGDTTGAGDGETPDPLLTLPLLRRSVPETGDAEHPGGFERALATAYAHGVPLRLAP
APDAASLAVAAELPTYAFQRTHYWLDAPAAPAALPAGLDDAGHPLLSAALDLPGGRGT
VWTGALSAATLPWAADHSVHGRTVLPGTALLDLALHAAPRVGELTFEAPLVLPEDGEV
RLRVVLAEPDASGVRELSVHSAGEDGGWTRHATAVLDTGTTTAGEPAGAPPAAWPPGD
AEPLDLAAEYERFADAGIGYGPAFRGLRSAWRDGDAILADVRLPGELAGEADRYGIHP
ALLDAALHTAAAALGGAHGMLPFTWNGVTLHARGAHAIRVRLTPAGPDAVAVTAVDPA
GRPVFTAASLTLRPVTTGQLTAAEAARAPLYRVRWTGLPDTGTARDHTWAVAGGPGDL
LPGETPHHPDLASALADTGTAPFRVLADLRGYGTATPRELASQALALVQQWADAAEAA
EGRLVLVTRRAVDIGDGVTDPAAATVWGLVRAAQSEHPGCFALLDTDDSPRSRQLLPR
VAGTAEQLALRDGTLLAPSLTRATLPAGARLPALDGTVLITGGTGSLGAEAARHLVTR
HGARRLLLTSRSGPQAPGAAELVAELAALGAHADVAACDVADRAALRALLDRVPAGHP
LTAVLHTAGVLDDGVLTAQTPQRLAAVLRPKADAVRNLHELTQGHALSAFILYSSAAG
VLGSAGQSGYAAANAYLDSFAVWRRSRGLPAVSLGWGPWDGGGMASGLGGTDTARLRR
SGIAPLSRAEGLAALDAALAAGGDDTAPAHLLPIRVDAVTLRGADTVPAVLRDLAGTA

Query Match 35.6%; Score 91.8; DB 1; Length 43280;

```
                 FEGTGASRIDLPTYAFQHEHLWAVPPAPEAVAAADPDDAAFWTAVEDGDVSALTAALG
                 TDEDSVAAVLPALTSWRRARRDRSTVDAWRYRVAWKPLGGTLPHPSLTGTWLLVTADG
                 IDDTDVAGALETYGAEVRRLVLDEECVDRAVLRERLAGAEDVTGIVSVLAAAERTDAV
                 PGTSLVLGTALTVALIQALGDAEIDAPVWALTRGAVSTGRADELTAPVQAQVTGIGWT
                 AALEHPQRWGGTLDLPAALDARAAQRLAAVLSGALGSDDQLAIRPSGVFTRRIVRAEA

  Query Match           36.2%;  Score 93.4;  DB 1;  Length 123580;
  Best Local Similarity   63.8%;  Pred. No. 2.1e-06;
  Matches  164;  Conservative    0;  Mismatches   81;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||| |||| ||| | || |||||| |||    |  ||| || |||||||| ||| || |
Db      50176 CTGGCCGCCCTGGACGCCGAGCAGCGGCATCAGGCGCTGCTCGACCTGGTCCGCACCGAG 50235

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              ||    ||||||||   | ||  |||    ||   |  | || | ||    | |
Db      50236 ACGGCCGCGGTGCTCGGCCACTCCGGGGCGGACGCCGTCCCGGCCGAGCGGGAC------ 50289

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                 |||  ||  || |||||||||||  | ||||  ||||||||||| |  || ||
Db      50290 ------TTCAACCGGCTGGGCTTCGACTCGCTGATGGCGGTCGAACTGCGGACGCGGCTG 50343

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
                ||  ||||| ||    ||| |||| |  |||||| |||||||||| | |||||||
Db      50344 GCCACCGCCACCGGAGCCCGGCTGCCGGCCACGCTCGTCTTCGACCACCCGACGCCGGAC 50403

Qy        241 GCGGTCGCCGCGCGCCT 257
              |||||||||  || |||
Db      50404 GCGGTCGCCCGGCACCT 50420


RESULT 7
AX211739
LOCUS       AX211739               125401 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 35 from Patent WO0159126.
ACCESSION   AX211739
VERSION     AX211739.1  GI:15523950
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 125401)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 35 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .125401
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT    15664 a  49692 c  42871 g  17174 t
ORIGIN

  Query Match           36.2%;  Score 93.4;  DB 6;  Length 125401;
  Best Local Similarity   63.8%;  Pred. No. 2.1e-06;
  Matches  164;  Conservative    0;  Mismatches   81;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||| |||| ||| | || |||||| |||    |  ||| || |||||||| ||| || |
Db      50584 CTGGCCGCCCTGGACGCCGAGCAGCGGCATCAGGCGCTGCTCGACCTGGTCCGCACCGAG 50643

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              ||    ||||||||   | ||  |||    ||   |  | || | ||    | |
```

```
Db      50644 ACGGCCGCGGTGCTCGGCCACTCCGGGGCGGACGCCGTCCCGGCCGAGCGGGAC------ 50697

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                 |||  ||  || |||||||||||  | ||||  ||||||||||| |  || ||
Db      50698 ------TTCAACCGGCTGGGCTTCGACTCGCTGATGGCGGTCGAACTGCGGACGCGGCTG 50751

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
                ||  ||||| ||    ||| |||| |  |||||| |||||||||| | |||||||
Db      50752 GCCACCGCCACCGGAGCCCGGCTGCCGGCCACGCTCGTCTTCGACCACCCGACGCCGGAC 50811

Qy        241 GCGGTCGCCGCGCGCCT 257
              |||||||||  || |||
Db      50812 GCGGTCGCCCGGCACCT 50828


RESULT 8
SFU78289
LOCUS       SFU78289               43280 bp    DNA     linear   BCT 13-AUG-1997
DEFINITION  Streptomyces fradiae tylactone synthase, starter module and modules
            1-7, (tylG) gene, complete cds.
ACCESSION   U78289
VERSION     U78289.1  GI:2317859
KEYWORDS    .
SOURCE      Streptomyces fradiae.
  ORGANISM  Streptomyces fradiae
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 43280)
  AUTHORS   DeHoff,B.S., Sutton,K.L. and Rosteck,P.R. Jr.
  TITLE     Sequence of Streptomyces fradiae tylactone synthase gene tylG
  JOURNAL   Unpublished (1996)
REFERENCE   2  (bases 1 to 43280)
  AUTHORS   DeHoff,B.S., Sutton,K.L. and Rosteck,P.R. Jr.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-1996) Eli Lilly and Company, Lilly Corporate
            Center, Indianapolis, IN 46285, USA
COMMENT     tylG.
FEATURES             Location/Qualifiers
     source          1. .43280
                     /organism="Streptomyces fradiae"
                     /db_xref="taxon:1906"
     gene            1. .43280
                     /gene="tylG"
     CDS             816. .14234
                     /gene="tylG"
                     /note="ORF1; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="tylactone synthase starter module and modules 1
                     & 2"
                     /protein_id="AAB66504.1"
                     /db_xref="GI:2317860"
                     /translation="MSSALRRAVQSNCGYGDLMTSNTAAQNTGDQEDVDGPDSTHGGE
                     IAVVGMSCRLPGAAGVEEFWELLRSGRGMPTRQDDGTWRAALEDHAGFDAGFFGMNAR
                     QAAATDPQHRLMLELGWEALEDAGIVPGDLTGTDTGVFAGVASDDYAVLTRRSAVSAG
                     GYTATGLHRALAANRLSHFLGLRGPSLVVDSAQSASLVAVQLACESLRRGETSLAVAG
                     GVNLILTEESTTVMERMGALSPDGRCHTFDARANGYVRGEGGGAVVLKPLDAALADGD
                     RVYCVIKGGAVNNDGGGASLTTPDREAQEAVLRQAYRRAGVSTGAVRYVELHGTGTRA
                     GDPVEAAALGAVLGAGADSGRSTPLAVGSVKTNVGHLEGAAGIVGLIKATLCVRKGEL
                     VPSLNFSTPNPDIPLDDLRLRVQTERQEWNEEDDRPRVAGVSSFGMGGTNVHLVIAEA
                     PAAAGSSGAGGSGAGSGAGISAVSGVVPVVVSGRSRVVVREAAGRLAEVVEAGGVGLA
                     DVAVTMADRSRFGYRAVVLARGEAELAGRLRALAGGDPDAGVVTGAVLDGGVVVGAAP
                     GGAGAAGGAGAAGGAGGGGVVLVFPGQGTQWVGMGAGLLGSSEVFAASMRECARALSV
                     HVGWDLLEVVSGGAGLERVDVVQPVTWAVMVSLARYWQAMGVDVAAVVGHSQGEIAAA
                     TVAGALSLEDAAAVVALRAGLIGRYLAGRGAMAAVPLPAGEVEAGLAKWPGVEVAAVN
                     GPASTVVSGDRRAVAGYVAVCQAEGVQARLIPVDYASHSRHVEDLKGELERVLSGIRP
                     RSPRVPVCSTVAGEQPGEPVFDAGYWFRNLRNRVEFSAVVGGLLEEGHRRFIEVSAHP
                     VLVHAIEQTAEAADRSVHATGTLRRQDDSPHRLLTSTAEAWAHGATLTWDPALPPGHL
                     TTLPTYPFNHHHYWLDTIDGGGGDDATQEKESGPLTRELRGLPSSQKQLGFLLDLVCR
                     HTAVVLGLDTAAEVDPDLSFKKQGIQSMTGVELRNRLLTETGLALPTTLVYDRPTPRA
                     LAQFLHTELLDGSPSGSVLAPAQKSFEAQEPIAVVGMGCRFPGGVGSPEALWRLVVEG
                     VDAVSPFPGDRGWDVEGLYDPEPGVAGKSYVREGGFLHDAAEFDAEFFGISPREAVAM
                     DPQQRLLLETSWEAIERAGIDPHSLHGSRTGVYAGVMPQEYGPRLAEGAEGSDGYLLT
```

```
                /gene="nysG"
CDS             complement(867. .2684)
                /gene="nysG"
                /function="presumably involved in efflux of nystatin"
                /note="putative transporter (ABC family)"
                /codon_start=1
                /transl_table=11
                /product="NysG"
                /protein_id="AAF71763.1"
                /db_xref="GI:8050837"
                /translation="MASPDDLEEERTAPRPARRLVGLLRPHRRSVALAVSMGVGGIVL
                NAFGPLLLGRVTDLIADGVLGGVPGPAPGIDFAAIGRLLLVLLALYVVASLFMLAQGR
                LVASAVWRTIHELRRDAREKLTRLPLRHFDRQPAGELLSRTTNDIDNLQQTLQQTLAE
                LITSIFSLLTMLVLMLVISPSLAVVMLLSVPVSALIAARISKRAQPHYAAQWSANGTL
                NAHVEEVCTGHALIKGFDRRAAAEERFDACNDAVYRAAAKAQFASGAMEPVMMFVANL
                GYVLVAVIGAWKVINGTLTLGDVQAPILYARQFSQPIVEIASVAGRLQSGIASAQRVF
                TLLDAPEQAPDPLRPGTPARAEGRVEFTDVSFRYSPDTPLIENLSLTVEPGSTVAIVG
                PTGAGKTTLGNLLMRFYEPDSGRILLDGTDTATMTRDDLRSRFGLVLQDTWLFGGTIA
                ENIAYGAPGACRADIEEAARATCADRFIRTLPQGYDTVLDDESGTVSAGEKQLLTVAR
                AFLARPAVLVLDEATSSVDTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIVVM
                DAGRIVEQGTHDQLLCAQGLYARLHAARTHTPTAGAAAG"
gene            complement(2662. .4416)
                /gene="nysH"
CDS             complement(2662. .4416)
                /gene="nysH"
                /function="presumably involved in efflux of nystatin"
                /note="putative transporter (ABC family)"
                /codon_start=1
                /transl_table=11
                /product="NysH"
                /protein_id="AAF71764.1"
                /db_xref="GI:8050838"
                /translation="MLLRLLRAQLRPYAWATAALVALQLVQILGTLLLPTLGAALIDQ
                GVVRGDGGRITELGVVMGVVALVQIAAALGAAALAARTATAMGRDLRSALFRRILDFS
                AREIGRFGTPSLLTRSVNDVQQVQNLAQTGFGIVVCAPLMCLGSVLLALRQDVPLALL
                LVALVLVVAVCFGLLLARMGTLYARMQLTLDRLGRLLREAITGVRVVRSFVRDDHERA
                RFAQTNDAFLVVSRRVGRLIATMLPVVLLLMNGFTVALLWTGSHRIDAGRMPIGSLSA
                LLSYLSLILMSVVMLAFVFLSVPRARVCAGRIAEVLDTGSSVAPPAAPQPVRGPAGRI
                ELCAAGYRYPGAEEPVLRDVDLTVEPGERIAVLGSTGSGKTTLLNLVLRLADATEGAV
                RVGGTDVRELTAATLAAAVGFVPQRPYLFSGTVASNLRFGRPDATDEELWEALRVAQA
                ADFVARMPDGLDAEITQGGGNVSGGQRQRLSLARALLRRPEIYLFDDCFSALDQATDA
                ALRTALVPYTAGATVITVAQRISAGRDADRIVVLDRGRVVAQGTHDVLLRTSPTYREI
                ALSQLTEEEAAHGLAGRP"
gene            4714. .5748
                /gene="nysDIII"
CDS             4714. .5748
                /gene="nysDIII"
                /function="presumably involved in mycosamine biosynthesis"
                /note="putative dGDP-mannose-4,6-dehydratase"
                /codon_start=1
                /transl_table=11
                /product="NysDIII"
                /protein_id="AAF71765.1"
                /db_xref="GI:8050839"
                /translation="MSKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKSRV
                SRLASELDFIDGDLMDQGSLVSAVDTVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                VLRMLEAIRMVSGLSTSRTVSPRGQIRFYQASSSEMFGKAAETPQRETTLFHPRSPYG
                AAKAYGHYITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKQGLQDKL
                ALGNLDAVRDWGYAGDYVRAMHLMLQQDAGDDYVIGTGQMHSVRDAVRIAFEHVGLNW
                EDYVVIDPDLVRPAEVEVLCADSAKAQDRLGWKPDVDFPTLMRMMVDSDLAQVSRENQ
                YGDVLLAANW"
gene            5930. .34363
                /gene="nysI"
CDS             5930. .34363
                /gene="nysI"
                /function="responsible for condensation steps 9 to 14 in
                the nystatin polyketide backbone synthesis"
                /note="polyketide synthase"
                /codon_start=1
                /transl_table=11
                /product="NysI"
                /protein_id="AAF71766.1"
                /db_xref="GI:8050840"
                /translation="MDNEQKLRDYLKLATADLRRTRRRVHKLESAAQEPVAIIGMTCR
                YPGGVRSPEDLWRMVEAGEHGVTPFPTDRGWDLEALAAAPTASGGFLHDAPDFDADFF
                GISPREAVAMDPQQRVVLESAWEAFERAGIDPTSVKGSRTGVFIGAMAQDYRVGPADG
                AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                AGGVTIMSGPGTFIEMGRQGGLSADGRCRSFGDTADGTGWAEGVGILVLERLSDAVRN
                GHEILAVVRGTAVNQDGASNGLTAPNGPSQQQVIQQALVNARLAAGDIDVVEAHGTGT
                TLGDPVEAQALLATYGQNRPADRPLLLGSVKSNLSHTQAAAGVAGVIKMVMAMRHGTL
                PRTLHAEEPTHHVDWSQGAVRLLTDTTDWPATGAPRRAAVSSFGISGTNAHTIIEQAP
                EPQPEDAATAQDDAAGSTPATAPVVPGVVPVLLSGRTPDALRGQAAALRAALDTGRRP
                DLLDLAHSLATTRAGFEHRAVLLATDHPALTDGLTALADADDPAAAPAWITGTTRAET
                RLAVLFTGQGAQRLGAGRELAARFPAFATALDAALDAFTPHLDRPLREVLWGTDAALL
                DRTAYAQPALFAVEVALYRLIESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAATLVA
                ARGRLMQALPDGGAMIAVQASEADVAPLLAGHEDQVAIAAVNGPSAVVLSGAEATVTA
                LAEQLAADGRKTRRLRVSHAFHSPLMEPMLDAFRAVVEDLTLQPPLLPVVSNLTGKPA
                TVAQLTSADYWVDHVRHAVRFADGIDWLARHDTTAFLELGPDGVLSAMAQDCLDAADA
                DAVTLPALRAGRPEEHTLTTALAGLHVHGATLDWTGCFAGTGARRTDLPTYAFQRRRY
                WPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRG
                TTLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRT
                VNVHARPDTGDDTPWTAHATGVLTTADASRQLPASSEQGGTPLAGDPHPALDAAQWPP
                AGAEPLPLDGHYDRLADGGFGYGPVFQGLRAAWRGGDVVYAEVELPEAGRSDAEAFGL
                HPALLDAALHAAPFTGLGERGRGGLPFSWEGVSLHAGGATTLRVRLTPVADDALALTV
                ADGTGAPVLSVDSLVLRSVATQQLDTAAAVARDALFRLDWTPVQPTATDPGPVALLGA
                DPFGLLTHAGFADAPAYPDLAALAAADGPVPTTVVLSLAGTGDDAADPARSAHRCAAE
                ALAAVQTWLDHHERFAAARLVFVTRGATVGRDVAAAAVWGLVRSAQSENPGCFALVDL
                DPDGAVGAAALVAALVSGEPQLAVRGDVLRVARLVRRPLTEVGAGADGTGDGVGDGSG
                VSFSGEGAVLVTGGTGGLGAVLARHLVAEYGVRDLLLVSRSGERAVGAGELVAELAGV
                GARVRVVACDVTDRAAVVELVGGHAVSAVVHAAGVLDDGMVGALTGERLSAVLRPKVD
                AVWHLHEATRGLDLDAFVVFSSLAGVFGSPGQANYAAANAFLDALMTRRRAEGLPGLS
                LAWGPWEQSGGMTGTLTDVDAERLARSGVPPLSVAQGLALFDAAVAGTDATCVPVRLD
                LPVLRARGEVPPLLRSLIRVRARRAAVAGSATAGNLAQRLRRLDEDGRDEMVLDLVRG
                QVALVLGHATGGDVDAGRAFRDLGFDSLTAVELRNRLNTVTGLRLPATLVFDYPTVRH
                LATYVLDELLGTDAEVATVQPAAVAVADDPIVIVGMACRYPGGVSSPEDLWRVLTEGT
                DAVSGFPTNRGWDVESLYHPDPDHPGTSYTRSGGFLHEAGEFDPGFFGMSPREALATD
                SQQRLLLESSWEAIERAGIDPVSLRGSRTGVFAGVMYSDYSAMLASPEFEGFQGSGSS
                PSLASGRVAYTLGLEGPAVTVDTACSSSLVAMHWAMQALRSGECGLALAGGVTVMSTP
                AVFVDFARQRGLSPDGRCKAFADAADGVGWSEGVGVLVLERQSDAVRNGHEILAVVRG
                SAVNQDGASNGLTAPNGPSQQRVIRQALASGGLTAGDVDVVEAHGTGTTLGDPIEAQA
                LLATYGRDREPERPLLLGSVKSNLGHTQAAAGVAGVIKMVLAMRHGVVPRTLHVDAPS
                SHVDWSEGAVELLSEQAAWPETGRVRRAGVSSFGISGTNVHVIVEQAPGAKAIAAAGA
                ARRTPGAVPVLLSGRGRSALRGQAARLLGHLQARPDAELVDVALSLATTRSRFEQRAA
                VVAQDRDQLIASLGALAADRPDPAVVEGEAAGRGRTAVLFTGQGSQRAAMGRELHEVQ
                PEFAAAFDAVCAVFDPLLDRPLREVVFAEDGSDEAALLDETGWTQPALFAVEVALFRL
                VESWGVRPDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARATLMQALPTGGAMIAIQA
                TEDEIAAHLDDTVAIAAVNGPQSVVISGDEEAAETIAATFAERGRKTKRLRVSHAFHS
                PRMDGMLDAFRIVAEGLTYRAPRIPLVSDLTGRRADDAEVCTAEYWVRHVREAVRFAD
                CVRTLRDAGATTFLELGSDGLLTAMAEDTLGDDHDAELVPMLRAGRABELAAATALAR
                LQVRGVDVDWAAYLAGTGARRTDLPTYAFQHAYYWPQLPTPAAALAAADPADQQLWAA
                VERGDARELADILGLGEQDLTPLDSLLPALTSWRRGNQEKHLLDTLRYRVEWTRLSKP
                TAPVLDGTWLLVASDATAADQPALLDGLADALGSHGARVRRLLLDDSCADRAVLAERL
                ARTADVDAATQVLSVLPLDERDADDCPPLTRGLALTVALVQALADTGAQGRLWTATRG
                AVSTNPADPVTHPVQAAAWGLGRGVALEHPRLWGGLVDLPQVFDERAGQRLAGILAVK
                DAPDGEDQVALRATGVSGRRLVRHTVEALPTAAEFTATGTVLITGGTGGLGAEVARWL
                ARAGAQHLVLTSRRGPDAPGAAELRAELEGYGPSVSVVACDVADRDALAAVLTALPEE
                LPLTGVVHTAGVGHYGPLDTLSTAEFAGLTAAKLAGAAHLDALLADRELDFFVLFGSI
                AGVWGSGNQSAYGAANAYLDALALHRRARGLAATSVAWGPWAEAGMAADDAVSETLRR
                QGLGLLDPAPAMTELRRAVVRQDVTVTVADVDWQRYAPLFTSARPSALIAGLPEVRAL
                AADERTEQDATGASEVVTRVRALAEPEQLRLLTDLVRTESATVLGHSSADAVPEGRAF
                RDVGFDSLTAVELRKRLGAATGLSLPSTMVFDYPTPLELAQYLRAEILGAVLEVAGPV
                ATGGADDEPIAIIGMACRFPGGVSSPEQLWDLVASGTDAISEFPVNRGWQTGHLFDPD
                PDRPGTTYSTQGGFLHEADEFDPTFFGISPREALVMDPQQRLLLETTWESFERAGIRP
                ETLRSTLTGTFVGSSYQEYGLGAGDGTEGHMVTGSSPSVLSGRLSYVFGLEGPAVTVD
                TACSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNPFIAFSRQRALAKDGRCKAFS
                DDADGMTLAEGVGVVLVERLSDAQRNGHPVLAVLRGSAINQDGASNGLTAPNGPSQQR
                VIRQALANARLAPGDIDALEAHGTGTPLGDPIEAQALFATYGRDRDPESALLLGSVKS
                NIGHTQSAAGIASVIKMVMALRHSELPPTLHADAPSSHVDWSAGTVRLLTQARAWPET
                GRPRRAAVSSFGISGTNAHVLLEQAPVADTPAEERPAVAPVPIAAGVVPWVVTARSAA
                ALRGQAERLLAHAETVGTALPAAGPLDIGLSLVSARARFEHRAVVVPPAGTDPLAALR
                AVATDGPSPVVARGVADVEGRTVFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAAL
                AEFTDWSLVDVLRGVVGAPSLERVDVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQG
                EIAAAVVSGALSLRDGARVVALRSQAIGRALAGRGGMMSVALSVDVLEPRLVEFEGRV
                SVAAVNGPRSVVVAGEPEALDALHARLTADDIRARRIAVDYASHSHQVEDLHEELLEV
                LAELAPRTSEVPFFSTVTGDWLDTARMDAGYWFRNLRGRVRFADAVADLLAAEYRAFV
                EVSSHPVLSMAVQEAIDEAGVPAVAAGTLRRDQGGTDRFLLSAAEVFVRGVDVDWAGL
```

```
                EELSLDCLRLAGQSATTEQERAAVVMARVKIGWEIDPRLVSPWLDELVAALRRGDLRS
                GDAAWIVKHLAWQDRFPEAVESLSTLAAAGDAGEPGVEAELHFVRHWLRHSCPPLEAS
                VPGPSNPTRPRTPRLRRVSPASYAGEVLGKVLTEGPDDHTVAMAEEILRGCRYSETVV
                EALEAALLALVYAGRPDRAAFWCDGLLQQAGTRSTGTSVAILSGIRAEIALRQGALPA
                VEQYARSALSAISLPGWGVAVGSPLAALVHSTTASGMSGVAGAWLNQAVPDGMFRTRH
                GLLYTHARGHYHLAVDRPAAALDDFLTCGGLAKEWGMDVPTFLPWRTSAAQAQLALGN
                PGQARALVREQLGRPGGSTARARGVALRLLAATGEPGRRPALLRESVNLLESCGDQVE
                LLRALADRCQALDHVGAPAKARMTARHARNVADACGGDALFRKLFRDEVPDREGDPDG
                LTGEEQGSASLTAAERRVTALAALGHSNREIGRKLFITKSTVEQHLTRVYRKLGVRNR
                ADLGDLLAGTNLGAGTPAMAMASSAVAG"
CDS             10933. .11271
                /codon_start=1
                /transl_table=11
                /product="P450-like protein"
                /protein_id="BAB69191.1"
                /db_xref="GI:15823974"
                /translation="MEVGGQTIRAGEGVLCTLSTANRDEGAFTDPDELDLRRDARSHL
                SFGFGIHQCIGQSLARAEPQIALETLLRRLPGLRLAASFEELRFRDGVGFYGVKELPD
                PCPALPPDRP"
gene            11458. .29898
                /gene="olmA1"
CDS             11458. .29898
                /gene="olmA1"
                /codon_start=1
                /transl_table=11
                /product="modular polyketide synthase"
                /protein_id="BAB69192.1"
                /db_xref="GI:15823975"
                /translation="MACRLPGAAGPAEFLALLRNGEDAVGAAPGDRRYAPRRGGFLDS
                VDEFDAGFFGISPREAAAMDPQQRLVLELSWEALEDAGIPPARLGGSGAGVFVGAIAD
                DYATLSRAAGADAATPQTTTGLNRGIIANRVSYTLGLRGPSFTVDSGQSSSLVAVHLA
                AESLRRGECSVALAGGVNLILAEDSTVAVERFGALSPDGYCYTFDARANGYVRGEGGG
                VVVLKPLADAIADGDDIHCILAGSAVNNDGGGEGLTVPDRAGQEAVLTAAYEQAGISP
                AAVGYVELHGTGTPTGDPVEAAAVGAVLGAGRPAGQPLLVGSVKTNIGHLEGAAGIAG
                LLKAVLAVRHREIPATLHHASPNPAIPMAELGIGVNTALTPWPEQAGPLVVGVSSFGM
                GGTNCHVVLTEPAGYDAAARPATAGIRPAGTPAPLLVSGRDEQALRAQARKLRDHLDT
                HPGLPLRDVAHTLAAARAPLARRAVLLAADREDVSAGLARLADGTPGADLVRGTVSPG
                ALAFVFTGQGSQRPGMTAELYRTSAEYAAALDEVCASLDPWLRRPLRDVLFAAEGSDE
                AALLDRTEFTQPALFAVEVALFRYAEHCGLTPRFLLGHSVGELAAAHVAGVLSLADAC
                TLVAARGRLMQDRPATGAMVSVQATEDELAPFVSESVTLAALNGPASTVIAGDEDAVL
                AVAAHWAARGRKTKRLRVSHAFHSPHMDGMLTEFRRIAAGLTFGAPRIPIVSNETGTL
                LTEADARSPEYWVRHARAAVRFLDGVRLLEDQGVTTLLELGPDGVLSAMARDCLRDAG
                TVCVPLLRGRDRTEPEALLAALAGLYVRGVPMDWERTATEAGARRVPLPTYAFQRRRH
                WLPEPSASASRPVVNGGGHAGQAGHAGQAGRAGQAGHAGHDGEPAPAVRDRTTPELVR
                DRTTLELVRDATAQVLGHASPDDVATHEAFRELGFSSLMLAELGERLTEATGRRVPTT
                LLFDHPTPDALARRLEADGAERTDAGPVSPPAAAAHDDDPIVIVGMGCRLPGGVRSPE
                DLWRFVAAGGDAISALPVDRGWEFAADFPGEGGFLADVAEFDAGFFGISPREALAMDP
                QQRLLLETSWEALERAGVDALKLRGSRTGVFVGASPSEYGPRLHETSDADGHVLTGTA
                PSVLSGRIAYVLGLEGPALTVDTACSSSLVALHLAAQALRGGECDLALAGGVTVMATS
                GMFAEFARQGGLAPDGRCKAFADGADGTGWGEGVGMLAVQRLSDAVRDGRPVLAVVRG
                SAVNSDGASNGLTAPNGPSQQRVIRQALASAGLSAADVDMVEAHGTGTSLGDPIEAQA
                LLATYGQDRPDDRPLWLGSVKSNIGHTQAAAGVAGVIKSVLALRNGLLPKTLHVDEPT
                PEVDWSAGAVELLTEGREWPETDGPRRAGVSAFGISGTNAHVILEQAPSVEEPVTAPA
                DDGLLVPWVVSARSEEALRAQARRLADHVRESGLRPADVGLSLAVTRAGLEHRAVLVA
                EDEAGFLAHLDALAEGRTPAGVVVEGGTAFLFSGQGSQRAGMGRELYSAHPVFAAAFD
                EVCAAFDGLLPGALREVVFGGAGLDRTEWAQPALFALEVALFELVSSWGVRADVLLGH
                SIGELAAAYVAGVWSLADACRIVAARGRLMQALPEGGAMVAVEAAEDELPDLPGGVSV
                AAVNGPRSLVLSGDEEPVTALAQAFAEQGRRTKQLTVSHAFHSARMEPMLAEFAETLA
                AVEFRPPRIPVVSNVTGEMAGDEFTTPAYWVRHVREAVRFADGVGAAHAAGVRRYLEL
                GPDGTLTTLAHDTAHDAGRDVAAVPALRGDQPETRTVVTAVGRLYATGLTVDWPAFFA
                PYGARPTELPTYAFQHRRYWLEPVAPSSAPPANPLRYRTAWASVPDGPRPSLSGTWAV
                VVPASGSYEDIVAGVTAALRDHGARPVVVEVDPAGPERADVADTLRAALGEDAGGVVS
                LLGTAGSLSSGVAATAVLVQAWTDLGGGARLWCLTRGAVSVSPSDPLAAPAQAQLWGM
                GRVAALEHPERWGGLVDLPAVPDDRAWTRLCAVLAGGSGEDQVAVRAAGLFARRLGRA
                GTRADAAPAGPSPWRTDGTVLVTGGTGALGAHLAQWLAAAGAAHVLLTSRRGPDADGA
                AELTARLRETGTEVTVAACDVTDRDALAELLSGLPADRPLTGVVHAAGVLDDGVLDSL
                TPDRFAAVAGPKVTGAWHLHELTRDLDLSAFVLFSSFAGTVGLAGQANYAAANAYLDA
                LAVHRAQLGLPATAVAWGSWSGAGMAADTEAARRQLARTGLVPLAPDAALAALGQVMD
                DGETAVTVADVDWDRFAAGFGPGRPSPLLTGVPEVRRIEERETGRAGAPAETDRLGGL

Query Match             36.8%;  Score 95;  DB 1;  Length 104326;
Best Local Similarity   64.2%;  Pred. No. 1.2e-06;
Matches  165;  Conservative    0;  Mismatches   80;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||| |||||||||||||| |  ||| |||   |||. | ||||| ||||||||||  |||
Db      29395 CTGGCCGCGCTGTCCCCCGCCGAGCGGCACCGGCTCGTACTGGATCTGGTGCGCGAGCAC 29454

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              |||  ||  || ||   |||           || | ||    ||   || |||
Db      29455 ACCGCGGGCGTCCTCGGTCACG-----------GCTCGGCCGACGACATCGACCCCGAC 29502

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                |   |||       |||||||||||||| |||  |  |||||||| |||||||| ||
Db      29503 CAGGCGTTCAAGACGCTCGGCTTCGACTCGCTCACCGCGGTCGAACTCCGCAACCGGCTG 29562

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
                | | ||||||   ||||    | |||| ||| ||  ||||||||||| |||| |||||
Db      29563 CGGACGGCCACCTCCCTGACCGTCCCCGCGACCCTGGTCTTCGACCACCCCACCCCGGCG 29622

Qy        241 GCGGTCGCCGCGCGCCT 257
              ||| || ||||||  ||
Db      29623 GCGCTCACCGCGCATCT 29639

RESULT 6
AF263912
LOCUS       AF263912               123580 bp    DNA     linear   BCT 24-MAY-2000
DEFINITION  Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF263912
VERSION     AF263912.1  GI:8050835
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., StrLm,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Biosynthesis of the polyene antifungal antibiotic nystatin in
            Streptomyces noursei ATCC 11455: analysis of the gene cluster and
            deduction of the biosynthetic pathway
  JOURNAL   Chem. Biol. 7 (6), 395-403 (2000)
  MEDLINE   20334850
   PUBMED   10873841
REFERENCE   2  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
            N-7489, Norway
FEATURES             Location/Qualifiers
     source          1. .123580
                     /organism="Streptomyces noursei"
                     /strain="ATCC 11455"
                     /db_xref="ATCC:11455"
                     /db_xref="taxon:1971"
     gene            complement(46. .783)
                     /gene="nysF"
     CDS             complement(46. .783)
                     /gene="nysF"
                     /function="presumably post-translationally modifies ACP
                     domains on PKS"
                     /note="putative 4'-phosphopantheteine transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysF"
                     /protein_id="AAF71762.1"
                     /db_xref="GI:8050836"
                     /translation="MIELILPATVATEAAYDDRPRPGDRLLSSEREVIARAVESRQRE
                     FTTVRHLARRALRRLGHPDRAILPNRRGAPQWPPGIVGSMTHCAGYRAAAVSPAELSA
                     AVSIDAEPNGPLPAGVLNAIALPSERPHLVALAAHRPDVHWDRLLFSAKESVFKAWYP
                     LTQRELDFSEAEIVIDPTQGAFTARLLVPGPLLGGRRVTVFPGRWHSTPALLTTAVHL
                     PAPTPRRDREHRTHLTVNSPLPRPTFG"
     gene            complement(867. .2684)
```

```
Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1378 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 1437

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db       1438 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTAGACCACACCACGCCGGCC 1497

Qy        241 GCGGTCGCCGCGCGCCTT 258
              ||||||||||||||||||
Db       1498 GCGGTCGCCGCGCGCCTT 1515


RESULT 5
AB070940
LOCUS       AB070940              104326 bp    DNA     linear   BCT 22-JAN-2002
DEFINITION  Streptomyces avermitilis oligomycin biosynthetic gene cluster.
ACCESSION   AB070940
VERSION     AB070940.1  GI:15823967
KEYWORDS    .
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1
  AUTHORS   Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C.,
            Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T.,
            Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
  TITLE     Genome sequence of an industrial microorganism Streptomyces
            avermitilis: deducing the ability of producing secondary
            metabolites
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
  MEDLINE   21477403
REFERENCE   2  (bases 1 to 104326)
  AUTHORS   Ikeda,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-SEP-2001) Haruo Ikeda, Kitasato University, Dept. of
            Microbial Chemistry, School of Pharmaceutical Sciences; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .104326
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            complement(1. .1134)
                     /gene="mcmA"
     CDS             complement(1. .1134)
                     /gene="mcmA"
                     /note="methylmalonyl-CoA mutase"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB69185.1"
                     /db_xref="GI:15823968"
                     /translation="MTGYSTAADANARFRQLVGSGTRAVPVVFDLPTRAGLDSDAPAA
                     RGLVGRRGVAVDSIDDMRVLFGGLRLDRVVPSLAVDAPAAPLLVLYQLVAEEHGAPAQ
                     RLAGAVRGDVFKEFLARGPGVFPLRPSLRLAMNVFAYCQAELPRWQCLSVCGHHLAGT
                     GADPALEVAFTVAGGIEYLRAAVTAGLDVRTVAPRLTLCLVGGTTGRENRAKLRATRR
                     VWTRTANERFGSGRTAPRLYVRTPRDGLATLACTEDIEAGATALVDRIERLGGTLTVL
                     EHGLAGQLLGRPHGDRPATPRPHPPHQDSLDLQARQTERLAKLRAWRVQPAVDEALLR
                     LRKAAQGDDNVLYPMREALAARATVGEVCDALRSTWGTGPSAT"
     gene            complement(1131. .1886)
                     /gene="pptA"
     CDS             complement(1131. .1886)
                     /gene="pptA"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphopantheteinyl transferase"
                     /protein_id="BAB69186.1"
                     /db_xref="GI:15823969"
                     /translation="MRAHEVFADVPVSSLHPREAALVQDAVESRRREFATGRHCARRA
                     LASLGVPPGPLLATPRGAPRWPAGVCGAITHCAGYRAAAVARAPGIRAVGIDAEPDRP
                     LPEGVLEAVALPAEQFQVKEHTARHPGVVWGKLLFSAKESVYKAWFPATGRPLGFEEA
                     LIDLDPVSSTFRVRVLPTALRAAEPELRDLHGRWLVRHGIIVTAVVVDGPGRDTPHAP
                     SPYDRTKHSAPTGRSADRPGGRPTDRATDRPTA"
     gene            complement(1910. .2875)
                     /gene="simX4 homolog"
     CDS             complement(1910. .2875)
                     /gene="simX4 homolog"
                     /codon_start=1
                     /transl_table=11
                     /product="SimX4 homolog"
                     /protein_id="BAB69187.1"
                     /db_xref="GI:15823970"
                     /translation="MSQDRHDGSTAALGGPSHGGRLLAVSDIHVRHRQNRDIVQDLHP
                     GCAEDWLIVAGDVAESVADITWALSLLRDRFAKVIWAPGNHELWTAPGDPDELRGERK
                     YRHLVDLCRALGVVTPEDPYPVWHGAGGPVAVAPLFLLYDYSFLPPGAWSKSHALRLA
                     DDAGVICTDEFLLHPDPYPSREAWCRARVRATRARLDALPDDLPTVLVNHFPLVREPT
                     RVLRHPEFALWCGTTATADWHLRYRAAAVVHGHLHIPRTIRCDDVPHHEVSLGYPREW
                     QRYGRRPDWPVQILTPAPAPAPAPVPVPVPAPPGHRTASPADRRP"
     gene            3213. .3971
                     /gene="olmC"
     CDS             3213. .3971
                     /gene="olmC"
                     /codon_start=1
                     /transl_table=11
                     /product="thioesterase"
                     /protein_id="BAB69188.1"
                     /db_xref="GI:15823971"
                     /translation="MSANRVDGNLWVRQFHPAPTAATKLVCLPHAGGSASYFLPVSKA
                     LAPGVDVLAIQYPGRQDRRHEKCIDNIAELADAVTAVLEPLADRPLVLFGHSMGATLA
                     FEVALRLERKGIVPLALFASGRRAPSRHRDESVHLRGDDEIIAELKTLSGTRSEVLGD
                     EEILRMVLPAIRSDYKAAETYRYAGGARLTAPIHAHFGVEDPRVSLDEAQAWADHTTG
                     HFELRSHPGGHFYLNDQAPRVIEEIARVSSSVAS"
     gene            4643. .7438
                     /gene="olmRI"
     CDS             4643. .7438
                     /gene="olmRI"
                     /codon_start=1
                     /transl_table=11
                     /product="luxR-family transcriptional regulatory protein"
                     /protein_id="BAB69189.1"
                     /db_xref="GI:15823972"
                     /translation="MLIGRGQEIDRLRAAYSSVRSTGRSRAVLVESAAGCGKSEILED
                     LVRHVESLGLIALRATAIAEERNTPLSVLRSLVNGAPFPDDLVRRFHTMSDAAARMIS
                     TEGRDKARRDAAWYELMRTFCTEVRALAQRAPLVVAVDDIHHGDADSLRQLLHLARHC
                     RNVPVLLTFTQSLAGGPADRTFGAELLRQPNLEHIRLPRFCRQDIADMLAEQGRADAG
                     QGTVERYFAVSGGNPLLLRALLEDARLEPDDTVDAQDVGPRAGQAYGRAVLACLDRGG
                     EKAHEVAAGLAVLGGAFTHELLSRLLDISAAALTENLRALRAAGIVVGTSFQHPAAKA
                     AVLDSLEPSHRLRLHQRSAALLHASGAPAALAAEHLHEAGQADGAWAVGVLQQAAEQA
                     LLADEARRAVSYLELAHAASADPRGRTEIKIRLGSVTRRVSAADAERHVDDALRGFGA
                     GLLTPSDTGGLAALLLAHGRLEQARSVLEQGRRPGDGDADRTAARRQLAEIRDSVLNR
                     GAWCAGTRAAWHENPAGRPSTGRAGTGLRAWLSGSHAALLAQRIAGTEGLLEITHLTD
                     ATFDLVLNAVWTLLALGAVERALHWCDTFLREAKRRDAPGWEAVFAAVRAEGALRGGN
                     LAEAERDARRCFAVLPERHRSALEGSAVSYLVTVQIARGKYEDAARQLSRPMADSLPD
                     TIHWLDYLRARGLYYMATRQYHSALRDFHEVGRLAERWGTDWPALLPWRTDTAEALLR
                     LGEREQGLRLATEQLAMVQDGAPRIKGISLRLQGLASEPAARLGLLNRAVVELSRADD
                     RLELSRTLFDLADAHRETRQTRQAAMVLRRARQLAKSCGAQPPRDRIGSPDAAGASGE
                     LFGEGPERNADWGHLSESERRVAGLAASGYTNREISVRLYITMSTVEQHLTRVYRKLK
                     INGREELLAEARAGVQV"
     gene            7658. .10486
                     /gene="olmRII"
     CDS             7658. .10486
                     /gene="olmRII"
                     /codon_start=1
                     /transl_table=11
                     /product="luxR-family transcriptional regulatory protein"
                     /protein_id="BAB69190.1"
                     /db_xref="GI:15823973"
                     /translation="MLASAGQGEGGVVLVNAGPGVGRTALLDAFLRKATESGARVCAA
                     AGSSAETEAEFGVVGQLFPPTGPLGAGFRLARAIGSRDAGPGLDADRLFDALCQELRR
                     RPTVIAVDDAQLADAASLRFLRYLVRRLRTAPVMMVLTEPTGPDALPPSFQAELLRHP
                     QCRRLGLLPLSRDGVTQVVSAQVGSDRVPRLAADFHAISGGNPLLVHALLADHRAAHR
                     LSPVPSVPTVPSVSSASSAASASSAASAASAVSAAPAASAAFAKAALAWAYRDEPVLF
                     DVACGFAVLGESATPARVACLVDRGADAVDQVMTALDTAGLLAGGTFRSALVGKALLE
                     HVDVETRAELRRRAARLLHGDGAPATVVARHLLAAPVAEPWAGRTLLAAAEKAMQEGR
```

```
                 IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
                 TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
                 PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
                 ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
                 YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
                 AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL

  Query Match           100.0%;  Score 258;  DB 1;  Length 64957;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-32;
  Matches  258;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1196 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 1255

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1256 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 1315

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1316 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 1375

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1376 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 1435

Qy        241 GCGGTCGCCGCGCGCCTT 258
              ||||||||||||||||||
Db       1436 GCGGTCGCCGCGCGCCTT 1453


RESULT 4
AF275943
LOCUS       AF275943               11096 bp    DNA     linear   BCT 02-SEP-2000
DEFINITION  Streptomyces avermitilis avermectin polyketide synthase gene,
            partial cds.
ACCESSION   AF275943
VERSION     AF275943.1  GI:9964075
KEYWORDS    .
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 11096)
  AUTHORS   Hong,Y.-S. and Lee,J.J.
  TITLE     Targeted Gene Disruption of the avermectin O-methyltransferase gene
            and polyketide synthase gene from Streptomyces avermitilis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 11096)
  AUTHORS   Hong,Y.-S. and Lee,J.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUN-2000) Anticancer Agent Research Laboratory, Korea
            Research Institute of Bioscience and Biotechnology, P.O. Box 116,
            YuSong-Gu, Taejon 305-600, South Korea
FEATURES             Location/Qualifiers
     source          1. .11096
                     /organism="Streptomyces avermitilis"
                     /strain="ATCC31271"
                     /db_xref="ATCC:31271"
                     /db_xref="taxon:33903"
     CDS             220. .>11096
                     /note="Avr1"
                     /codon_start=1
                     /transl_table=11
                     /product="avermectin polyketide synthase"
                     /protein_id="AAG09812.1"
                     /db_xref="GI:9964076"
                     /translation="MADEADGGVVFVFPGQGPQWPGMGRELLDASDVFRESVRACEAE
                     FAPYVDWSVEQVLRDSPDAPGLDRVDVVQRTLFAVMISLAALWRSQGVEPCAVLGHSL
                     GEIAAAHVSGGLSLADAARVGDAWSQAQTTLAGTGALVSVAATPDELLPRIAPWTEDN
                     PARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTRMIPVDVPAHSPLMYAIEER
                     VVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRNMSSTVRFEPAARLLLQQGP
                     KTFVEMSPHPVLTMGLQELAADLGDTTGTADTVIMGTLRRGQGTLDHFLTSLAQLRGH
                     GETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLNDDGNERTASDAGPSASFAHL
                     GFDSVMGVELRNRLSKATGLRLPVTLILDHTTPAAVAARLRTAALGHLDEDTAPVPDS
                     PSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAASGGDAIGPFPTDRGWPTEQR
                     HAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALAMDPQQRLLLETSWEAFERA
                     GIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVEGHILTGTTGSVLSGRSAYS
                     FGLEGPAITVDTGCSASLVTLHLACQSLRSGECTFALAGGVSDVHPGMFIEFSRQCGL
                     SVDGRCKAVSAAADGTGWGEGIGLLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQAL
                     ANAGLSVADVDVVEGHGTGTTLGDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMA
                     AAGVGGVIKMVMALREGVLPRTLHVDEPSPQGLAAGAVRLLTEAVPWPGDAAGRLRRA
                     GVSSFGIGGTNAHVILEEAPAAGGCVAGGRVLEGAPGLAISVAESVAAPVAVSAPVAE
                     SVPVPVPVPVPVPVSARSEAGLRAQAEALRQYEAVQPDVSLADVGAGLACRQAVLEHH
                     VVILAACTSSSRSAARTTARSSSTARPQARPAPTSARRIGALAAGSGSAALTTGHAPG
                     GDRGGVVFVFPWQGGQWAGMGVRLLCLLRVFARRMQACEEALAPWVDWSVVDILRRDA
                     GDAVWEQADVVQPVLFSVMVSLAALWRSYGIEPNEVLGHSKDEIAAAHIYGALSLKDA
                     AKTVALPPQEVEQLIGERGGRLWVAAVNGPRSTAVSGDAEAVVEVLAYCAGTGVRPRI
                     PVDYASHCPHVQPLREELLELLGDISPQPYGVPFFSTVEGTWLDTTTLDAAYWYRNLH
                     QPVRFSHDVQALADQGHRVLLEVSPHPTLVPAIEDTTEDTARRRHCDRQPPPRRERHP
                     LLPQRLRLDPYYRHRQTHHVAPLLDPPRHLPPPLDAPRPAHLPFQHQHYWLESSQPGA
                     GSGSGAGAGSGAGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVV
                     PALSAWHRHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNIL
                     TNLHHHGITPIPLPLTTHHTNPQHLHHTLHHTRQQAQNHTTGAITGLLSLLALDETPH
                     PHHPHTPTGTLLNLTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLAR
                     TTLLEHPTHTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTL
                     TPTHQPPTPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLT
                     TQLQQKGIHLTITTCDTSTPRPTHNNSLNTIPPQHPVTTVIHTGGILDDATLTNLTPT
                     QLNNVLRAKAHSAHLLHQLTQHTPLTAFVLYSSAAATFGAPGQANYAAANAYLDALAH
                     HRHTHHLPATSIAWGTWQGNGLADSDKARAYLDRRGFRPMSPELATAAVTQAIADTER
                     PYVVIADIDWSKIEHTSQTSDLVSAARERE PAVQRPTPPAELHKTLAHQTSADQRAAL
                     LELVRDHVAAVLRHADPKAIAPDQSFRALGFDSLTAVEFRNLLIKATGLRLPVSLVFD
                     HPTPAKLAVHLQNQLRGTAAESAPSAAAVTAEASVTEPIAIVGMACRFPGGVTSADDF
                     WDLISSEQDAIGGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGIS
                     PREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDYALRVHNAGQS
                     TDGFALTGTAGSVISGRISYTFGFEGPAVSVDTACSSSLVALHLACQALRAGECSMAL
                     AGGVTVMSSPGAFVEFSRQRGLAADGHCKAFSAAADGTGWGEGVGMLLVERLSDAHRN
                     GHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAGLSAGDVDAVEAHGTGT
                     TLGDPIEAQALLATYGQDRAGEGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLL
                     PRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEA
                     PADDVPGGPPAGEGDAGSDDEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPG
                     LDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGT
                     GEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQN
                     DNDNEDAAALLQQTRYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITGAHLAG
                     ILTLTDATTLITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSL
                     VISGTPHTVQHITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHT
                     PLITANTPPDQLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHH
                     NLPNTPTTTLTLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHAQPTPHPHPLDLPTYP
                     FQHHHYWLESTQPGAGNVSAAGLDPTEHPLLGATLELATDGGALLAGRLSLRSHPWLA
                     DHAVGGTVLLSGATFLELALHAGTYVGCDRVDELTLHAPLVVPVDGGVSVQVGVAAAD
                     GKGRRLVSVYARGGSACGGGASGGVWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAV
                     AVDVDGVRDRLAGAGCVLGPVFSGLRAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPAL
                     LNGVVQPLSVLLPGGTGFGKGAGFGKGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGG
                     GREAVSVVVGDEAGVPVASVDRLELRPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVP
                     VCGQAWAWHEDVGESGGGPVPGVVVLRCPDAGAGGGGGGGGGGGVGEVVGGVLGVVQG
                     WLGLERFAGSRLVVVTRGAVVAGPEDGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDT
                     DTGTDLDTGAGAGWGVDGGRVAAVVACGEPQLAVRGERLLAARLTRLESSGDVPAQRS
                     GDTRARRSDVPAQRSGGVPARRSVDVSGREVLPWLSGGSVLVTGGTGVLGAAVARHLA
                     GVCGVRDLLLVSRRGPDAPGAEGLRRSWPRGAEVRIVACDVGERREVVRLAGGCSCRV
                     SVFVDFSAVASVTVRLPVASDVRKEAAMAYATVEEFTDYLLDPDP"
BASE COUNT     1756 a   4059 c   3528 g   1753 t
ORIGIN

  Query Match           99.4%;  Score 256.4;  DB 1;  Length 11096;
  Best Local Similarity  99.6%;  Pred. No. 3.8e-32;
  Matches  257;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1258 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 1317

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1318 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 1377
```

RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
GDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAG
LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
THHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGAL
ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
TAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSA
ARAREREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
MTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
NLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY
VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGGASGG
VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
LRCPDAGAGGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPE
DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
VSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
SLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLGLS"

gene 12071. .30790
/gene="aveA2"

CDS 12071. .30790
/gene="aveA2"
/note="contains Ave module 3, module 4, module 5 and module 6
multifunctional polyketide synthase"
/codon_start=1
/transl_table=11
/product="type I polyketide synthase AVES 2"
/protein_id="BAA84475.1"
/db_xref="GI:5902892"
/translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGLFN
EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
ACQALRSGECTMALAGGASVMATPFVFTEFSRQRGLAADGRCKAFSAAADGTGWSEGV
GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
NDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALHRL
LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
TQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
NRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLTAVPADEPIAIV
GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
LTYHDYAARFPTAPAGFEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
HLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSAAADGTGWGE
GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAD
LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAGDDAGTGGEATGPA
AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH
GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
QGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
PQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
QHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG
TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
GTWQGNGLATGQVSEHLRRRGMFAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA
AVLGRDDSEAIPPGRLFRDLGFDSLAAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ
FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI
GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAAL
LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
TLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
KATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD

```
  JOURNAL   Patent: WO 0001827-A 1 13-JAN-2000;
            KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
            STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PFIZER (US); MCARTHUR
            HAMISH ALASTAIR IRVIN (US); STUTZMAN ENGWALL KIM JONELLE (US)
FEATURES             Location/Qualifiers
     source          1. .12381
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     1884 a   4561 c   4005 g   1931 t
ORIGIN

  Query Match           100.0%;  Score 258;  DB 6;  Length 12381;
  Best Local Similarity  100.0%;  Pred. No. 2e-32;
  Matches  258;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1913 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 1972

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1973 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 2032

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2033 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 2092

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2093 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 2152

Qy        241 GCGGTCGCCGCGCGCCTT 258
              ||||||||||||||||||
Db       2153 GCGGTCGCCGCGCGCCTT 2170


RESULT 2
E38020
LOCUS       E38020                 30690 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Avermectin aglycon synthase gene.
ACCESSION   E38020
VERSION     E38020.1  GI:18626909
KEYWORDS    JP 2000245457-A/1.
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 30690)
  AUTHORS   Omura,S. and Ikeda,H.
  TITLE     Avermectin aglycon synthase gene
  JOURNAL   Patent: JP 2000245457-A 1 12-SEP-2000;
            THE KITASATO INSTITUTE
COMMENT     OS   Streptomyces avermitilis
            PN   JP 2000245457-A/1
            PD   12-SEP-2000
            PF   24-FEB-1999 JP 1999046961
            PR
            PI   SATOSHI OMURA,HARUO IKEDA
            PC   C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC   (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(11916)
            FT   CDS             (11971). .(30687).
FEATURES             Location/Qualifiers
     source          1. .30690
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     5355 a  12455 c   8617 g   4263 t
ORIGIN

  Query Match           100.0%;  Score 258;  DB 6;  Length 30690;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-32;
  Matches  258;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 1155

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 1215

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 1275

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1276 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 1335

Qy        241 GCGGTCGCCGCGCGCCTT 258
              ||||||||||||||||||
Db       1336 GCGGTCGCCGCGCGCCTT 1353


RESULT 3
AB032367
LOCUS       AB032367               64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
            Sciences, Kitasato University, Microbial chemistry; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .64957
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            101. .12019
                     /gene="aveA1"
     CDS             101. .12019
                     /gene="aveA1"
                     /note="contains a loading module, Ave module 1 and module
                     2
                     multifunctional polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="type I polyketide synthase AVES 1"
                     /protein_id="BAA84474.1"
                     /db_xref="GI:5902891"
                     /translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
                     ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
                     AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
                     VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTR
                     MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
                     MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 500.468 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_1096_1353
Perfect score: 258
Sequence: 1 ctgaccgcgctgtcccccac.....ccgcggtcgccgcgcgcctt 258

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		258	100.0	12381	6	AX006889	AX006889 Sequence
2		258	100.0	30690	6	E38020	E38020 Avermectin
3		258	100.0	64957	1	AB032367	AB032367 Streptomy
4		256.4	99.4	11096	1	AF275943	AF275943 Streptomy
5		95	36.8	104326	1	AB070940	AB070940 Streptomy
6		93.4	36.2	123580	1	AF263912	AF263912 Streptomy
7		93.4	36.2	125401	6	AX211739	AX211739 Sequence
8		91.8	35.6	43280	1	SFU78289	U78289 Streptomyce
9	c	91.8	35.6	104326	1	AB070940	AB070940 Streptomy
10		90.2	35.0	53784	1	AMM223012	AJ223012 Amycolato
11		90.2	35.0	53789	6	A69720	A69720 Sequence 3
12		90.2	35.0	90445	1	AF040570	AF040570 Amycolato
13	c	87	33.7	30000	6	AX250263	AX250263 Sequence
14		85.4	33.1	5676	6	A69718	A69718 Sequence 1
15		85	32.9	78210	1	AB070949	AB070949 Streptomy
16		84.6	32.8	32870	1	AF007101	AF007101 Streptomy
17		83.8	32.5	20394	1	SNA132222	AJ132222 Streptomy
18		83.8	32.5	20394	6	AX067996	AX067996 Sequence
19	c	83.8	32.5	84985	1	SNA278573	AJ278573 Streptomy
20	c	83.2	32.2	77534	1	AF235504	AF235504 Streptomy
21		82.6	32.0	2700	6	AX211733	AX211733 Sequence
22		82.6	32.0	2706	1	AF420594	AF420594 Streptomy
23		82.6	32.0	65140	6	AX211705	AX211705 Sequence
24		82.4	31.9	5544	1	SNA132221	AJ132221 Streptomy
25		82.4	31.9	5544	6	AX067994	AX067994 Sequence
26		82.4	31.9	84985	1	SNA278573	AJ278573 Streptomy
27		82.4	31.9	113193	1	AF357202	AF357202 Streptomy
28		81.6	31.6	30000	6	AX250263	AX250263 Sequence
29		80.6	31.2	41097	1	AF016585	AF016585 Streptomy
30	c	79.8	30.9	82746	1	AF453501	AF453501 Actinosyn
31		79.4	30.8	37948	1	AF079138	AF079138 Streptomy
32	c	79	30.6	24225	1	SC2C4	AL512902 Streptomy
33		79	30.6	28732	1	AF082100	AF082100 Streptomy
34	c	77.6	30.1	26195	1	SC1G7	AL591083 Streptomy
35		77.4	30.0	1681	6	AR173228	AR173228 Sequence
36		77.4	30.0	11219	1	SERERYAA	M63676 Saccharopol
37		77.4	30.0	11219	6	AR049367	AR049367 Sequence
38		77.4	30.0	11219	6	AR095528	AR095528 Sequence
39		77.4	30.0	31422	6	E38021	E38021 Avermectin
40	c	77.4	30.0	64957	1	AB032367	AB032367 Streptomy
41		75	29.1	758	1	AF144051	AF144051 Streptomy
42	c	74.6	28.9	27522	1	AB070942	AB070942 Streptomy
43		74.6	28.9	30000	6	AX250262	AX250262 Sequence
44		74.4	28.8	107379	1	SHGCPIR	X86780 S.hygroscop
45	c	74.2	28.8	107379	1	SHGCPIR	X86780 S.hygroscop

ALIGNMENTS

RESULT 1
AX006889
LOCUS AX006889 12381 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 1 from Patent WO0001827.
ACCESSION AX006889
VERSION AX006889.1 GI:9994904
KEYWORDS .
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 12381)
AUTHORS Kellenberger,J.L., Leadlay,P.F., Staunton,J., Mcarthur,H.A. and
Stutzman-Engwall,K.J.
TITLE Polyketides, their preparation, and materials for use therein

```
Qy          61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
               |  |||||||||     |||              ||  | |  | |    |||| |  |
Db       19536 GTCGCGGCGGTGCTCGGACACG-----------GCGACCCCGCCGCGATCGGCGCCGCC 19583

Qy         121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                 |  |||    ||  ||| ||||||||| |||  | ||||||  || ||||||||  ||
Db       19584 CGCACCTTCAAGGACGCCGGATTCGACTCCCTCACCGCTGTCGACCTCCGCAACCGGCTG 19643

Qy         181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
               | ||    ||| || |||||| ||||||.||| ||| ||||||||||| |||| |||  |
Db       19644 AACACACGCACCGGACTGCGGCTGCCCGCGACCCTCGTCTTCGACCACCCCACACCGCTC 19703

Qy         241 GCGGTCGCCGCGCGCCT 257
               ||  ||||||  | |||
Db       19704 GCCCTCGCCGAACTCCT 19720
```

Search completed: June 17, 2003, 13:48:04
Job time : 57.414 secs

```
FT                   /note= "ORF5 encodes protein shown in AAW23720"
XX
PN   EP791656-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301066.
XX
PR   22-FEB-1996;   96US-0012050.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Burgett SG,  Kuhstoss SA,  Rao RN,  Richardson MA;
PI   Rosteck PR;
XX
DR   WPI; 1997-418047/39.
DR   P-PSDB; AAW23716-W23720.
XX
PT   DNA encoding Streptomyces ambofaciens platenolide synthase domain -
PT   for production of spiramycin-related polyketide antibiotics
XX
PS   Claim 9; Pages 8-33; 81pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. The DNA can be used to produce compounds
CC   exhibiting antibiotic activity based on the platenolide structure,
CC   including specifically the macrolide antibiotic spiramycin and spiramycin
CC   analogues and derivatives. Modifications of the platenolide synthase DNA
CC   sequence can be made so as to change the number and type of carboxylic
CC   acids incorporated into the growing polyketide chain and to change the
CC   kind of post-condensation processing that is conducted.
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

  Query Match             31.9%;  Score 82.2;  DB 18;  Length 44377;
  Best Local Similarity   61.1%;  Pred. No. 8.6e-09;
  Matches  157;  Conservative    0;  Mismatches   88;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||  | |  |  ||| |  |  |||    | | || || || ||||||||||  || |
Db      19476 CTCGCGGGACGTTCCGCACCCGAGCGTACGGCTCTCCTGCTCGACCTGGTGCGGACCGAG 19535

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
                |  |||||||||     |||           ||  | |  | |    |||| |  |
Db      19536 GTCGCGGCGGTGCTCGGACACG-----------GCGACCCCGCCGCGATCGGCGCCGCC 19583

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                |  |||    ||  ||| ||||||||| |||  | ||||| || |||||||| ||
Db      19584 CGCACCTTCAAGGACGCCGGATTCGACTCCCTCACCGCTGTCGACCTCCGCAACCGGCTG 19643

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              | ||    ||| || |||||| |||||| ||| ||| |||||||||| |||| |||  |
Db      19644 AACACACGCACCGGACTGCGGCTGCCCGCGACCCTCGTCTTCGACCACCCCACACCGCTC 19703

Qy        241 GCGGTCGCCGCGCGCCT 257
              ||  ||||||  | |||
Db      19704 GCCCTCGCCGAACTCCT 19720


RESULT 15
AAT80414
ID   AAT80414 standard; DNA; 44377 BP.
XX
AC   AAT80414;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Platenolide synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   platenolide synthase gene cluster; platenolide production; srmG gene;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   CDS             350..14002
FT                   /*tag=  a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in AAW22606"
FT   CDS             14046..20036
FT                   /*tag=  b
FT                   /note= "ORF2 encodes protein shown in AAW22607"
FT   CDS             20110..31284
FT                   /*tag=  c
FT                   /transl_except= (pos:20111..20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in AAW22608"
FT   CDS             31329..36071
FT                   /*tag=  d
FT                   /note= "ORF4 encodes protein shown in AAW22609"
FT   CDS             36155..41830
FT                   /*tag=  e
FT                   /note= "ORF5 encodes protein shown in AAW22610"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22606-W22610.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Example 2; Pages 110-134; 220pp; English.
XX
CC   This sequence represents the platenolide synthase gene cluster of the
CC   invention. This sequence is referred to as the srmG gene, and was
CC   isolated from Streptomyces ambofaciens. This sequence encodes the
CC   multi-functional proteins which direct the synthesis of the polyketide
CC   platenolide. Platenolide is the basic building block of the macrolide
CC   antibiotic spiramycin. This sequence was used along with the tylG gene
CC   (see AAT80413) to create a hybrid ORF1 sequence (see AAT80415). The tylG
CC   gene is the tylactone synthase gene cluster of the invention. The tylG
CC   sequence was isolated from Streptomyces fradiae, and encodes
CC   multifunctional proteins which direct the synthesis of the polyketide
CC   tylactone. Tylactone is the basic building block of the antibiotic
CC   tylosin. The hybrid sequence can be used to transform S. ambofaciens
CC   lacking the srmG ORF1 sequence, or S. fradiae lacking the tylG ORF1
CC   sequence, so that they can produce polyketides. The DNA sequence can be
CC   modified so as to alter the type of carboxylic acids incorporated, the
CC   number of carboxylic acids incorporated and/or the post-condensation
CC   reactions performed, thereby resulting in novel tylosin-related
CC   polyketides.
XX
SQ   Sequence 44377 BP; 4965 A; 15552 C; 17381 G; 6479 T; 0 other;

  Query Match             31.9%;  Score 82.2;  DB 18;  Length 44377;
  Best Local Similarity   61.1%;  Pred. No. 8.6e-09;
  Matches  157;  Conservative    0;  Mismatches   88;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||  | |  |  ||| |  |  |||    | | || || || ||||||||||  || |
Db      19476 CTCGCGGGACGTTCCGCACCCGAGCGTACGGCTCTCCTGCTCGACCTGGTGCGGACCGAG 19535
```

```
XX
PS   Claim 2; Page 116-151; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

  Query Match             32.0%;  Score 82.6;  DB 22;  Length 65140;
  Best Local Similarity   57.6%;  Pred. No. 7e-09;
  Matches  148;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||  ||| || |  | ||  ||  ||||   |   ||  | ||||    | | ||| | |
Db      33083 CTCGCCGAGCAGCTCACCCGGCTCCAGCGCGCCGAGCGACGGGACACCCTCCTCGCGCTC 33142

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
                |   |   |         | |    | |  | | |     |  | || ||| ||
Db      33143 GTCCGCGACCAGGCCGCGATGGTCCTCGGCCACACCTCGGGCGACGGCGTCGACCCGTCC 33202

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                 |||  | |||||||||||||||||  |||  |  |||||||| ||||||||| ||
Db      33203 CGCGCCTTCCGCGACCTCGGCTTCGACTCGCTCACCGCGGTCGAACTCCGCAACCGCATC 33262

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              ||  |||||| |||||||| | || |  |||  | |||||||| || ||||  | |
Db      33263 GGCGCGGCCACCGGCCTGCGGCTACCGGCCACGGCCGTCTTCGACTACCCCACCGCCGAT 33322

Qy        241 GCGGTCGCCGCGCGCCT 257
              ||  ||||||| | |||
Db      33323 GCCCTCGCCGCACACCT 33339


RESULT 13
AAF24891
ID   AAF24891 standard; DNA; 5544 BP.
XX
AC   AAF24891;
XX
DT   20-APR-2001  (first entry)
XX
DE   Pimaricin biosynthesis associated polyketide synthase gene.
XX
KW   Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW   anticancer; pimaricin; ss.
XX
OS   Streptomyces natalensis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..5544
FT                   /*tag= a
FT                   /product= "polyketide synthase"
XX
PN   WO200077222-A1.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-EP06227.
XX
PR   14-JUN-1999;   99EP-0201893.
XX
PA   (STAM ) DSM NV.
XX
PI   Martin JF,  Aparicio JF,  Colina AJ;
XX
DR   WPI; 2001-080693/09.
DR   P-PSDB; AAB31557.
XX
PT   New polynucleotides encoding enzymes involved in the biosynthesis of
PT   pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT   the synthesis of new compounds -
XX
PS   Disclosure; Page 39-46; 116pp; English.
XX
CC   The present sequence encodes a polyketide synthase which is associated
CC   with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC   is useful for the oxidative modification of a methyl group of a suitable
CC   compound, e.g. a bioactive compound including a secondary metabolite,
CC   antibiotics and anticancer agents. Recombinant cells comprising the
CC   gene are useful for the production of pimaricin. The polyketide synthase
CC   polynucleotide may be over expressed in Streptomyces, leading to an
CC   increase in the biosynthesis of pimaricin, as a source of primers for
CC   amplification reaction and as probes.
XX
SQ   Sequence 5544 BP; 699 A; 2052 C; 2023 G; 770 T; 0 other;

  Query Match             31.9%;  Score 82.4;  DB 22;  Length 5544;
  Best Local Similarity   60.0%;  Pred. No. 7.9e-09;
  Matches  156;  Conservative    0;  Mismatches  101;  Indels    3;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||| |||  ||   |   | | |||||  ||||  | |||| |||||| | |||| |
Db       5035 CTGTCCGGTCTCGACGAGGCCGAACAGCACGCGCTCGTACTGGCCCTGGTCCTCGCCGAG 5094

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCA--- 117
              |||  ||| |||||   || | | |  |       ||   | | || | | ||
Db       5095 ACCTCGGCCGTGCTCGGCGGCCAGGAGACGCCTGGCGAGGAGCCGCACGGCGAGGAGGGC 5154

Qy        118 TCGGCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGC 177
                  |  |||    |  | ||| || ||||  | |  |  ||||||||||||||||||
Db       5155 CACCGCACGTTCAAGGAGATGGGCATCAACTCGTTGAACGCCGTCGAACTGCGCAACCGC 5214

Qy        178 CTCAGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCG 237
              || | |   ||||||| ||||||| | ||||  |||||| ||| |||| || ||||||||
Db       5215 CTTATCGCAGCCACGGACCTGCGGCTTCCCGCCACGCTCGTCTACGACTACCCCACGCCG 5274

Qy        238 GCCGCGGTCGCCGCGCGCCT 257
                || ||||  |   | | |
Db       5275 AAGGCCGTCGTCCGACTCGT 5294


RESULT 14
AAT78508
ID   AAT78508 standard; DNA; 44377 BP.
XX
AC   AAT78508;
XX
DT   26-FEB-1998  (first entry)
XX
DE   Platenolide synthase gene cluster.
XX
KW   Platenolide synthase gene cluster; platenolide production; srmG gene;
KW   multi-functional protein; macrolide antibiotic; spiramycin; ss.
XX
OS   Streptomyces ambofaciens.
XX
FH   Key             Location/Qualifiers
FT   CDS             350..14002
FT                   /*tag= a
FT                   /transl_except= (pos:350..352, aa:Met)
FT                   /note= "ORF1 encodes protein shown in AAW23716"
FT   CDS             14046..20036
FT                   /*tag= b
FT                   /note= "ORF2 encodes protein shown in AAW23717"
FT   CDS             20110..31284
FT                   /*tag= c
FT                   /transl_except= (pos:20111..20113, aa:Met)
FT                   /note= "ORF3 encodes protein shown in AAW23718"
FT   CDS             31329..36071
FT                   /*tag= d
FT                   /note= "ORF4 encodes protein shown in AAW23719"
FT   CDS             36155..41830
FT                   /*tag= e
```

```
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10153.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Example 2; Page 186-187; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a ERD48 insert DNA in pKR48 which is used in an
CC   exemplification of the invention.
XX
SQ   Sequence 2700 BP; 324 A; 1087 C; 876 G; 413 T; 0 other;

  Query Match            32.0%;  Score 82.6;  DB 22;  Length 2700;
  Best Local Similarity  57.6%;  Pred. No. 7.2e-09;
  Matches  148;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||  ||| || |  | ||  ||  ||||    |   ||  | ||||    | | ||| | |
Db       1962 CTCGCCGAGCAGCTCACCCGGCTCCAGCGCGCCGAGCGACGGGACACCCTCCTCGCGCTC 2021

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
                |   |  |        | |    | |   | | |     |  | || ||| ||
Db       2022 GTCCGCGACCAGGCCGCGATGGTCCTCGGCCACACCTCGGGCGACGGCGTCGACCCGTCC 2081

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                 |||  | ||||||||||||||||  |||  |  ||||||| |||||||| ||
Db       2082 CGCGCCTTCCGCGACCTCGGCTTCGACTCGCTCACCGCGGTCGAACTCCGCAACCGCATC 2141

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
               ||  |||||| |||||||| | || |  |||  | |||||||| || ||||  | |
Db       2142 GGCGCGGCCACCGGCCTGCGGCTACCGGCCACGGCCGTCTTCGACTACCCCACCGCCGAT 2201

Qy        241 GCGGTCGCCGCGCGCCT 257
               ||  ||||||| | |||
Db       2202 GCCCTCGCCGCACACCT 2218


RESULT 12
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..1035)
FT                   /*tag=  a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement (1056..2576)
FT                   /*tag=  b
FT                   /product= "NysD1 protein"
FT   CDS             2806..6906
FT                   /*tag=  c
FT                   /product= "NysA protein"
FT   CDS             6952..16530
FT                   /*tag=  d
FT                   /product= "NysB protein"
FT   CDS             16550..49840
FT                   /*tag=  e
FT                   /product= "NysC protein"
FT   CDS             50260..51015
FT                   /*tag=  f
FT                   /product= "NysE protein"
FT   CDS             51405..54305
FT                   /*tag=  g
FT                   /product= "NysR1 protein"
FT   CDS             54329..57190
FT                   /*tag=  h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180..59963
FT                   /*tag=  i
FT                   /product= "NysR3 protein"
FT   CDS             60415..61047
FT                   /*tag=  j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736..62497
FT                   /*tag=  k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (62551..63615)
FT                   /*tag=  l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765..64961
FT                   /*tag=  m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
```

```
FT                   /note= "KS4"
FT   misc_feature    complement (28750..28960)
FT                   /*tag= ad
FT                   /note= "ACP3"
FT   misc_feature    complement (29092..29740)
FT                   /*tag= ae
FT                   /note= "KR3"
FT   misc_feature    complement (29869..31018)
FT                   /*tag= af
FT                   /note= "DH3 (inactive)"
FT   misc_feature    complement (31018..32185)
FT                   /*tag= ag
FT                   /note= "AT3"
FT   misc_feature    complement (32185..33439)
FT                   /*tag= ah
FT                   /note= "KS3"
FT   misc_feature    complement (33505..33715)
FT                   /*tag= ai
FT                   /note= "ACP2"
FT   misc_feature    complement (33823..34480)
FT                   /*tag= aj
FT                   /note= "KR2"
FT   misc_feature    complement (34606..35749)
FT                   /*tag= ak
FT   misc_feature    complement (28750..28960)
FT                   /*tag= al
FT                   /note= "DH2 (inactive)"
FT   misc_feature    complement (35749..37144)
FT                   /*tag= am
FT                   /note= "AT2"
FT   misc_feature    complement (37145..38296)
FT                   /*tag= an
FT                   /note= "KS2"
FT   misc_feature    complement (38371..38581)
FT                   /*tag= ao
FT                   /note= "ACP1"
FT   misc_feature    complement (38677..39307)
FT                   /*tag= ap
FT                   /note= "KR1"
FT   misc_feature    complement (39442..40609)
FT                   /*tag= aq
FT                   /note= "DH1"
FT   misc_feature    complement (40609..41842)
FT                   /*tag= ar
FT                   /note= "AT1"
FT   misc_feature    complement (41842..43093)
FT                   /*tag= as
FT                   /note= "KS of extender module 1 (KS1)"
FT   misc_feature    complement (43144..43660)
FT                   /*tag= at
FT                   /note= "ACP of loading domain"
FT   misc_feature    complement (43777..44629)
FT                   /*tag= au
FT                   /note= "ER of loading domain"
FT   misc_feature    complement (44974..46573)
FT                   /*tag= av
FT                   /note= "CoA ligase of laoding domain"
FT   CDS             46754..47788
FT                   /*tag= aw
FT                   /note= "fkbO gene"
FT   CDS             47785..52272
FT                   /*tag= ax
FT                   /note= "fkbP gene"
FT   CDS             52275..71465
FT                   /*tag= ay
FT                   /note= "fkbA gene"
FT   misc_feature    52362..53576
FT                   /*tag= az
FT                   /note= "KS7"
FT   misc_feature    53577..54716
FT                   /*tag= bb
FT                   /note= "AT7"
FT   misc_feature    54717..55871
FT                   /*tag= ba
FT                   /note= "DH7"
FT   misc_feature    56019..56819
FT                   /*tag= bc
FT                   /note= "ER7"
FT   misc_feature    56943..57575
FT                   /*tag= bd
FT                   /note= "KR7"
FT   misc_feature    57710..57920
FT                   /*tag= be
FT                   /note= "ACP7"
FT   misc_feature    57990..59243
FT                   /*tag= bf
FT                   /note= "KS8"
FT   misc_feature    59244..60398
FT                   /*tag= bg
FT                   /note= "AT8"
FT   misc_feature    60399..61412
FT                   /*tag= bh
FT                   /note= "DH8 (inactive)".

  Query Match             32.2%;  Score 83.2;  DB 21;  Length 77536;
  Best Local Similarity   62.5%;  Pred. No. 5.2e-09;
  Matches  130;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy          50 TGCGCGCCCACACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATG 109
               ||| ||| |       ||||    | || ||       | |   | ||| || ||
Db       38580 TGCTGGCCATCGTGTGTGCGGCCACGGCCGCCGTGCTCGGCCACGCCGACGCCTCCGAGA 38521

Qy         110 CCGGCCCATCGGCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGC 169
                |  ||  || ||   |||    |||||||| |||||||  |||  ||||||     |||
Db       38520 TCACGCCCACGACGGCGTTCAAGGACCTCGGCATCGACTCGCTCAGCGGTGTCCGGTTGC 38461

Qy         170 GCAACCGCCTCAGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACA 229
               ||||| |||||  | || | |||||  | ||| |  ||| |||| | ||||||||||
Db       38460 GCAACAGCCTCGCCGAGACGACGGGGGTACGGCTCTCCGCGACGGCCGTCTTCGACCACC 38401

Qy         230 CCACGCCGGCCGCGGTCGCCGCGCGCCT 257
               | || ||||||||| ||||||| |||||
Db       38400 CGACACCGGCCGCGCTCGCCGCCCGCCT 38373


RESULT 11
AAD17193
ID   AAD17193 standard; DNA; 2700 BP.
XX
AC   AAD17193;
XX
DT   29-NOV-2001  (first entry)
XX
DE   ERD48 insert DNA in pKR48.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..2699
FT                   /*tag= a
FT                   /product= "Protein encoded by ERD48 insert DNA"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
```

```
CC   The present sequence encodes a polyketide synthase which is associated
CC   with the biosynthesis of pimaricin. The polyketide synthase polypeptide
CC   is useful for the oxidative modification of a methyl group of a suitable
CC   compound, e.g. a bioactive compound including a secondary metabolite,
CC   antibiotics and anticancer agents. Recombinant cells comprising the
CC   gene are useful for the production of pimaricin. The polyketide synthase
CC   polynucleotide may be over expressed in Streptomyces, leading to an
CC   increase in the biosynthesis of pimaricin, as a source of primers for
CC   amplification reaction and as probes.
XX
SQ   Sequence 20394 BP; 2611 A; 7486 C; 7397 G; 2900 T; 0 other;

  Query Match            32.5%;  Score 83.8;  DB 22;  Length 20394;
  Best Local Similarity   61.5%;  Pred. No. 4e-09;
  Matches  158;  Conservative    0;  Mismatches   87;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              |||| ||  | |    |   | ||| ||    | ||||||||||||||| || |||||
Db       9574 CTGAGCGGACGGGGAACGGTGGAGCGGCGCGAGGTGCTCCTGGACCTGGTACGGGCCCAG 9633

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              | |  ||  || |||  | |||        |||   | || |||     | |
Db       9634 ATCGCGGTCGTCCTGGGCCACGCGAACCCGGAGACGATCGAGTCCACCCGTGTC------ 9687

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                 |||    |||||||||||||||| | |  |  ||||||||  ||||||||||||
Db       9688 ------TTCCAGGACCTCGGCTTCGACTCCCTGACCGCGGTCGAACTCCGCAACCGCCTC 9741

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              | ||| || || ||||||||  |  | |  ||   | ||||||| || ||||| ||| |
Db       9742 AACAACGCGACCGGCCTGCGCCTTTCGGCCACCGCCGTCTTCGACTACCCCACGGCGGAC 9801

Qy        241 GCGGTCGCCGCGCGCCT 257
              ||| ||| ||    |||
Db       9802 GCGCTCGTCGACTTCCT 9818


RESULT 10
AAA14651/c
ID   AAA14651 standard; DNA; 77536 BP.
XX
AC   AAA14651;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Nucleotide sequence of the FK-520 biosynthetic gene cluster.
XX
KW   FK-520; polyketide synthase; PKS; gene cluster; immunosuppressant;
KW   Streptomyces hygroscopicus var. ascomyceticus; immunophilin;
KW   FK-506 binding protein; polyketide compound; transplant rejection;
KW   graft-versus-host disease; uveitis; alopecia universalis;
KW   autoimmune chronic active hepatitis; inflammatory bowel disease;
KW   multiple sclerosis; primary biliary cirrhosis; scleroderma;
KW   neurite outgrowth; nerve regrowth; Parkinson's disease;
KW   Alzheimer's disease; stroke; traumatic spinal cord; brain injury;
KW   peripheral neuropathy; ss.
XX
OS   Streptomyces hygroscopicus.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (412..1836)
FT                   /*tag=  a
FT                   /note= "fkbW gene"
FT   CDS             complement (2020..3579)
FT                   /*tag=  b
FT                   /note= "fkbV gene"
FT   CDS             3969..4496
FT                   /*tag=  c
FT                   /note= "fkbR2 gene"
FT   CDS             complement (4595..5488)
FT                   /*tag=  d
FT                   /note= "fkbR1 gene"
FT   CDS             5601..6818
FT                   /*tag=  e
FT                   /note= "fkbE gene"
FT   CDS             6808..8052
FT                   /*tag=  f
FT                   /note= "fkbF gene"
FT   CDS             8156..8824
FT                   /*tag=  g
FT                   /note= "fkbG gene"
FT   CDS             complement (9122..9883)
FT                   /*tag=  h
FT                   /note= "fkbH gene"
FT   CDS             complement (9894..10994)
FT                   /*tag=  i
FT                   /note= "fkbI gene"
FT   CDS             complement (10987..11247)
FT                   /*tag=  j
FT                   /note= "fkbJ gene"
FT   CDS             complement (11244..12092)
FT                   /*tag=  k
FT                   /note= "fkbK gene"
FT   CDS             complement (12113..13150)
FT                   /*tag=  l
FT                   /note= "fkbL gene"
FT   CDS             complement (13212..23988)
FT                   /*tag=  m
FT                   /note= "fkbC gene"
FT   misc_feature    complement (13452..13662)
FT                   /*tag=  n
FT                   /note= "ACP6"
FT   misc_feature    complement (13761..14394)
FT                   /*tag=  o
FT                   /note= "KR6"
FT   misc_feature    complement (14517..15294)
FT                   /*tag=  p
FT                   /note= "ER6"
FT   misc_feature    complement (15438..16587)
FT                   /*tag=  q
FT                   /note= "dehydratase domain (DH) 6"
FT   misc_feature    complement (16587..17820)
FT                   /*tag=  r
FT                   /note= "acyltransferase domain (AT) 6"
FT   misc_feature    complement (17820..19053)
FT                   /*tag=  s
FT                   /note= "KS6"
FT   misc_feature    complement (19116..19326)
FT                   /*tag=  t
FT                   /note= "ACP5"
FT   misc_feature    complement (19464..20097)
FT                   /*tag=  u
FT                   /note= "KR5"
FT   misc_feature    complement (20241..21420)
FT                   /*tag=  v
FT                   /note= "DH5"
FT   misc_feature    complement (21420..22653)
FT                   /*tag=  w
FT                   /note= "AT5"
FT   misc_feature    complement (22653..23892)
FT                   /*tag=  x
FT                   /note= "KS5"
FT   CDS             complement (23992..46573)
FT                   /*tag=  y
FT                   /note= "fkbB gene"
FT   misc_feature    complement (24163..24373)
FT                   /*tag=  z
FT                   /note= "ACP4"
FT   misc_feature    complement (24997..26146)
FT                   /*tag=  aa
FT                   /note= "DH4 (inactive)"
FT   misc_feature    complement (26146..27430)
FT                   /*tag=  ab
FT                   /note= "AT4"
FT   misc_feature    complement (27430..28684)
FT                   /*tag=  ac
```

```
Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              |  ||| |||||     ||| ||              ||||   |  | | ||
Db      25291 GTCGCGGCCGTGCTCGGGCACGCCGA-----------CGCGAGCGCCGTCCGCGTCGAC 25338

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              ||   |||    ||  |||||||||||||  | |  |  ||||| |||||||||||| |
Db      25339 ACGGCCTTCAAGGACGCCGGCTTCGACTCGCTGACCGCGGTCGAGCTGCGCAACCGCATG 25398

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              | |  ||||| ||||||  | |||||| |||||||| |||||||| || | |  ||
Db      25399 CGGACCGCCACCGGCCTGAAGCTGCCCGCGACGCTCGTCTTCGACTACCCGAACCCCCAG 25458

Qy        241 GCGGTCGCCGCGCGCCT 257
              |||  |||||  || |||
Db      25459 GCGCTCGCCCGGCACCT 25475
```

RESULT 8
AAV21186
ID AAV21186 standard; DNA; 5676 BP.
XX
AC AAV21186;
XX
DT 24-JUL-1998 (first entry)
XX
DE Amycolatopsis mediterranei strain wt3136 5.7 kb KpnI DNA fragment.
XX
KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
KW polyketide synthase; actinomycete; ansamycin; ds.
XX
OS Amycolatopsis mediterranei.
XX
FH Key Location/Qualifiers
FT CDS 3..5676
FT /*tag= a
FT /product= "polyketide synthase"
FT /note= "no stop codon given"
XX
PN WO9807868-A1.
XX
PD 26-FEB-1998.
XX
PF 18-AUG-1997; 97WO-EP04495.
XX
PR 20-AUG-1996; 96EP-0810551.
XX
PA (NOVS) NOVARTIS AG.
XX
PI Engel N, Schupp T, Toupet C;
XX
DR WPI; 1998-169172/15.
DR P-PSDB; AAW52844.
XX
PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT to produce rifamycin and rifamycin analogues
XX
PS Claim 4; Page 38-43; 205pp; English.
XX
CC The present sequence represents Amycolatopsis mediterranei strain wt3136
CC 5.7 kb KpnI fragment DNA, from the present invention. The present
CC invention describes a Amycolatopsis mediterranei rifamycin synthesis
CC gene cluster DNA fragment comprising a DNA region involved directly or
CC indirectly in the gene cluster responsible for rifamycin synthesis,
CC including the adjacent DNA regions to the right and left which, by
CC reason of their function in connection with rifamycin biosynthesis,
CC qualify as constituents of this rifamycin gene cluster, and functional
CC fragments, derivatives or constituents of these. The Amycolatopsis
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC for producing rifamycin, rifamycin analogues or precursors. It can also
CC be used for inactivating or modifying genes involved in ansamycin or
CC rifamycin biosynthesis. The DNA can be used for constructing mutant
CC actinomycetes strains from which the natural rifamycin or ansamycin
CC biosynthesis gene cluster has been partly or completely deleted. The
CC DNA fragment can be used for assembling a library of polyketide
CC synthases, which can be used for assembling a library of polyketides.
CC A hybridisation probe of the invention can be used for identifying DNA
CC fragments involved in the biosynthesis of ansamycins.
XX
SQ Sequence 5676 BP; 674 A; 1990 C; 2205 G; 807 T; 0 other;

Query Match 33.1%; Score 85.4; DB 19; Length 5676;
Best Local Similarity 61.9%; Pred. No. 1.8e-09;
Matches 159; Conservative 0; Mismatches 86; Indels 12; Gaps 1;

```
Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||  ||| |||  | ||  || ||||| || ||||||| || ||| | || ||| | ||
Db       3072 CTCGCCGGGCTCGCGCCGGCGGAGCAGGAGGCGCTGCTGCTCGACGTCGTCCGCACGCAG 3131

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              |  |  ||||||     ||| |||     |||  |  |    ||||  | |
Db       3132 GTCGCGCTGGTGCTCGGGCACGCCGGGCCGGAGGCCGTCCGCGCGGACACGGCG------ 3185

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                 |||    ||  |||||||||||||  | | |  || ||||||||||||||| ||
Db       3186 ------TTCAAGGACACCGGCTTCGACTCGCTGACGTCGGTGGAACTGCGCAACCGGCTG 3239

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              || |||| |  || |||  | |||||| ||||||| |||||||| || | ||||||| |
Db       3240 CGCGAGGCGAGCGGGCTGAAGCTGCCCGCGACGCTCGTCTTCGACTACCCGACGCCGGTC 3299

Qy        241 GCGGTCGCCGCGCGCCT 257
              ||| | |||     |||
Db       3300 GCGCTGGCCCGCTACCT 3316
```

RESULT 9
AAF24892
ID AAF24892 standard; DNA; 20394 BP.
XX
AC AAF24892;
XX
DT 20-APR-2001 (first entry)
XX
DE Pimaricin biosynthesis associated polyketide synthase gene.
XX
KW Polyketide synthase; oxidative modification; metabolite; antibiotic;
KW anticancer; pimaricin; ss.
XX
OS Streptomyces natalensis.
XX
FH Key Location/Qualifiers
FT CDS 1..20394
FT /*tag= a
FT /product= "polyketide synthase"
XX
PN WO200077222-A1.
XX
PD 21-DEC-2000.
XX
PF 14-JUN-2000; 2000WO-EP06227.
XX
PR 14-JUN-1999; 99EP-0201893.
XX
PA (STAM) DSM NV.
XX
PI Martin JF, Aparicio JF, Colina AJ;
XX
DR WPI; 2001-080693/09.
DR P-PSDB; AAB31558.
XX
PT New polynucleotides encoding enzymes involved in the biosynthesis of
PT pimaricin, useful for modifying the biosynthesis of pimaricin and in
PT the synthesis of new compounds -
XX
PS Disclosure; Page 53-80; 116pp; English.
XX

```
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22601-W22605.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Claim 2; Pages 8-66; 220pp; English.
XX
CC   This sequence represents the tylactone synthase gene cluster of the
CC   invention. This sequence is also referred to as the tylG gene, and was
CC   isolated from Streptomyces fradiae. This sequence encodes multifunctional
CC   proteins which direct the synthesis of the polyketide tylactone, isolated
CC   from Streptomyces fradiae. Tylactone is the basic building block of the
CC   antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC   type of carboxylic acids incorporated, the number of carboxylic acids
CC   incorporated and/or the post-condensation reactions performed, thereby
CC   resulting in novel tylosin-related polyketides.
XX
SQ   Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;

  Query Match             35.6%;  Score 91.8;  DB 18;  Length 43280;
  Best Local Similarity   63.4%;  Pred. No. 8.1e-11;
  Matches  163;  Conservative    0;  Mismatches   82;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||||||| |||  |  |  | |   |||  | ||  |||||| |  | | |||||||||
Db      30648 CTGACCGGGCTCACGGCACAGGAAGGGCACGCCCTCGTCCTGGCCGAGATCCGCGCCCAC 30707

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
               |   ||||||||||  | ||| |    ||||    | | |   |||    | |
Db      30708 GCGGCGGCGGTGCTGGGCCACGGCTCCGACGACTCGATCCCCGAGGACCGGGCC------ 30761

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                  |||    |||||||||||||||||  |||  |  || ||  |||||||||| ||
Db      30762 ------TTCAAGGACCTCGGCTTCGACTCGCTCACCGCCGTGGAGATGCGCAACCGGCTG 30815

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              |||  |||||| |||| ||| ||||||  || ||| |||||||||| | || |||| |
Db      30816 AGCGCGGCCACCGGCCTCCGGCTGCCCGCCACCCTCGTCTTCGACCACCCGACCCCGGGC 30875

Qy        241 GCGGTCGCCGCGCGCCT 257
              | | | ||||  | |||
Db      30876 GAGCTGGCCGGCCACCT 30892


RESULT 7
AAV21187
ID   AAV21187 standard; DNA; 53789 BP.
XX
AC   AAV21187;
XX
DT   24-JUL-1998  (first entry)
XX
DE   Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
XX
KW   Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
KW   polyketide synthase; actinomycete; ansamycin; ds.
XX
OS   Amycolatopsis mediterranei.
XX
FH   Key             Location/Qualifiers
FT   CDS             1825..15543
FT                   /*tag=  a
FT                   /label= ORF_A
FT                   /product= "polyketide synthase"
FT   CDS             15550..30759
FT                   /*tag=  b
FT                   /label= ORF_B
FT                   /product= "polyketide synthase"
FT   CDS             30895..36060
FT                   /*tag=  c
FT                   /label= ORF_C
FT                   /product= "polyketide synthase"
FT   CDS             36259..41325
FT                   /*tag=  d
FT                   /label= ORF_D
FT                   /product= "polyketide synthase"
FT   CDS             41373..51614
FT                   /*tag=  e
FT                   /label= ORF_E
FT                   /product= "polyketide synthase"
FT   CDS             51713..5293
FT                   /*tag=  f
FT                   /label= ORF_F
FT                   /product= "polyketide synthase"
XX
PN   WO9807868-A1.
XX
PD   26-FEB-1998.
XX
PF   18-AUG-1997;   97WO-EP04495.
XX
PR   20-AUG-1996;   96EP-0810551.
XX
PA   (NOVS ) NOVARTIS AG.
XX
PI   Engel N,  Schupp T,  Toupet C;
XX
DR   WPI; 1998-169172/15.
DR   P-PSDB; AAW52845-W52850.
XX
PT   Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT   to produce rifamycin and rifamycin analogues
XX
PS   Claim 4; Page 53-102; 205pp; English.
XX
CC   The present sequence represents a Amycolatopsis mediterranei rifamycin
CC   synthesis gene cluster DNA fragment from the present invention. The
CC   DNA fragment comprises a DNA region involved directly or indirectly
CC   in the gene cluster responsible for rifamycin synthesis, including
CC   the adjacent DNA regions to the right and left which, by reason of
CC   their function in connection with rifamycin biosynthesis, qualify
CC   as constituents of this rifamycin gene cluster, and functional
CC   fragments, derivatives or constituents of these. The Amycolatopsis
CC   mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC   for producing rifamycin, rifamycin analogues or precursors. It can also
CC   be used for inactivating or modifying genes involved in ansamycin or
CC   rifamycin biosynthesis. The DNA can be used for constructing mutant
CC   actinomycetes strains from which the natural rifamycin or ansamycin
CC   biosynthesis gene cluster has been partly or completely deleted. The
CC   DNA fragment can be used for assembling a library of polyketide
CC   synthases, which can be used for assembling a library of polyketides.
CC   A hybridisation probe of the invention can be used for identifying DNA
CC   fragments involved in the biosynthesis of ansamycins.
XX
SQ   Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;

  Query Match             35.0%;  Score 90.2;  DB 19;  Length 53789;
  Best Local Similarity   63.0%;  Pred. No. 1.8e-10;
  Matches  162;  Conservative    0;  Mismatches   83;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||| ||| |||  |||| || | ||||  ||||||||||||||||||| || || ||
Db      25231 CTGGCCGGGCTCGCCCCGGCGGACCAGCTCACGCTGCTCCTGGACCTGGTCCGGGCGCAG 25290
```

```
RESULT 5
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag= a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag= b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag= c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag= d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag= e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag= f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag= g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag= h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match           36.2%;  Score 93.4;  DB 22;  Length 125401;
  Best Local Similarity   63.8%;  Pred. No. 3.7e-11;
  Matches  164;  Conservative    0;  Mismatches   81;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||| |||| |||  | ||  ||||| |||    |  ||| || |||||||| ||| || |
Db      50584 CTGGCCGCCCTGGACGCCGAGCAGCGGCATCAGGCGCTGCTCGACCTGGTCCGCACCGAG 50643

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              ||    ||||||||  | ||  |||    ||   |  | || | ||     | |
Db      50644 ACGGCCGCGGTGCTCGGCCACTCCGGGGCGGACGCCGTCCCGGCCGAGCGGGAC------ 50697

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                 |||  ||  || ||||||||||  | ||||  ||||||||||  |  || ||
Db      50698 ------TTCAACCGGCTGGGCTTCGACTCGCTGATGGCGGTCGAACTGCGGACGCGGCTG 50751

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
               ||  ||||| ||    ||| |||| |  |||||| |||||||||| | |||||| |
Db      50752 GCCACCGCCACCGGAGCCCGGCTGCCGGCCACGCTCGTCTTCGACCACCCGACGCCGGAC 50811

Qy        241 GCGGTCGCCGCGCGCCT 257
              |||||||||  || |||
Db      50812 GCGGTCGCCCGGCACCT 50828


RESULT 6
AAT80413
ID   AAT80413 standard; DNA; 43280 BP.
XX
AC   AAT80413;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Tylactone synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             816..14243
FT                   /*tag= a
FT                   /transl_except= (pos: 816..818, aa: Met)
FT                   /note= "ORF1 encodes protein shown in AAW22601"
FT   CDS             14351..19945
FT                   /*tag= b
FT                   /transl_except= (pos: 14351..14353, aa: Met)
FT                   /note= "ORF2 encodes protein shown in AAW22602"
FT   CDS             20010..31199
FT                   /*tag= c
FT                   /transl_except= (pos: 20010..20012, aa: Met)
FT                   /note= "ORF3 encodes protein shown in AAW22603"
FT   CDS             31232..36067
FT                   /*tag= d
FT                   /note= "ORF4 encodes protein shown in AAW22604"
FT   CDS             36249..41774
FT                   /*tag= e
FT                   /note= "ORF5 encodes protein shown in AAW22605"
XX
PN   EP791655-A2.
```

```
FT   CDS             11971..30690
FT                   /*tag=  b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23749, AAB23750.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 2; Page 66-134; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;

  Query Match            100.0%;  Score 258;  DB 21;  Length 30690;
  Best Local Similarity   100.0%;  Pred. No. 7e-46;
  Matches  258;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 1155

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 1215

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 1275

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1276 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 1335

Qy        241 GCGGTCGCCGCGCGCCTT 258
              ||||||||||||||||||
Db       1336 GCGGTCGCCGCGCGCCTT 1353


RESULT 4
AAH79277
ID   AAH79277 standard; DNA; 30690 BP.
XX
AC   AAH79277;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 1.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag=  a
FT                   /product= "AAG65264"
FT                   /partial
FT   CDS             11971..30690
FT                   /*tag=  b
FT                   /product= "AAG65265"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65264, AAG65265.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Example 2; Page 58-123; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;

  Query Match            100.0%;  Score 258;  DB 22;  Length 30690;
  Best Local Similarity   100.0%;  Pred. No. 7e-46;
  Matches  258;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 1155

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 1215

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 1275

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1276 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 1335

Qy        241 GCGGTCGCCGCGCGCCTT 258
              ||||||||||||||||||
Db       1336 GCGGTCGCCGCGCGCCTT 1353
```

PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR P-PSDB; AAG65268.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 10; Page 149-167; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is modified version of a
CC fragment of the S. avermitilis genome.
XX
SQ Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;

Query Match 100.0%; Score 258; DB 22; Length 11916;
Best Local Similarity 100.0%; Pred. No. 7e-46;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 1155

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 1215

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1216 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 1275

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1276 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 1335

Qy        241 GCGGTCGCCGCGCGCCTT 258
              ||||||||||||||||||
Db       1336 GCGGTCGCCGCGCGCCTT 1353
```

RESULT 2
AAZ58381
ID AAZ58381 standard; DNA; 12381 BP.
XX
AC AAZ58381;
XX
DT 23-MAY-2000 (first entry)
XX
DE Streptomyces avermitilis avermectin polyketide synthase modules 1+2.
XX
KW Polyketide synthase; avermectin; insecticide; ss.
XX
OS Streptomyces avermitilis.
XX
PN WO200001827-A2.
XX
PD 13-JAN-2000.
XX
PF 06-JUL-1999; 99WO-GB02158.
XX
PR 06-JUL-1998; 98GB-0014622.
XX
PA (BIOT-) BIOTICA TECHNOLOGY LTD.
PA (PFIZ) PFIZER INC.
XX

PI Kellenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ;
PI McArthur HAI;
XX
DR WPI; 2000-182117/16.
XX
PT Mutated Type I polyketide synthase containing a polylinker site in an
PT extension module for replacement of a reductive loop sequence, for
PT producing polyketides, e.g. B1 avermectin -
XX
PS Disclosure; Fig 7a-f; 75pp; English.
XX
CC The present sequence is that of DNA encoding the first 2 modules
CC of the avermectin polyketide synthase (PKS) of Streptomyces
CC avermitilis. The invention relates to nucleic acids encoding a
CC Type I PKS such as avermectin in which a polylinker with multiple
CC restriction sites replaces or 1 more PKS genes encoding enzymes
CC associated with reduction. Novel PKS are provided in which in
CC which the reductive loop in a selected module of the Type I PKS is
CC replaced with the equivalent segment from the same or different
CC PKS gene cluster or by a mutated or synthetic segment. Vectors and
CC host cells, and methods for producing novel polyketides by
CC culturing host cells are claimed. The polyketides obtained are
CC useful as antibiotics and insecticides. Fermentation products
CC containing C22-C23 dihydroavermectin, ivermectin and B1
CC avermectins are claimed.
XX
SQ Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;

Query Match 100.0%; Score 258; DB 21; Length 12381;
Best Local Similarity 100.0%; Pred. No. 7e-46;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1913 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 1972

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1973 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 2032

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2033 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 2092

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2093 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 2152

Qy        241 GCGGTCGCCGCGCGCCTT 258
              ||||||||||||||||||
Db       2153 GCGGTCGCCGCGCGCCTT 2170
```

RESULT 3
AAA92301
ID AAA92301 standard; DNA; 30690 BP.
XX
AC AAA92301;
XX
DT 10-JAN-2001 (first entry)
XX
DE S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
XX
KW Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW agrochemical; ds.
XX
OS Streptomyces avermitilis.
XX
FH Key Location/Qualifiers
FT CDS 1..11919
FT /*tag= a
FT /note= "avermectin aglycon synthase protein"

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 51.3231 Seconds
(without alignments)
11320.736 Million cell updates/sec

Title: US-09-914-286-1_COPY_1096_1353
Perfect score: 258
Sequence: 1 ctgaccgcgctgtcccccac.....ccgcggtcgccgcgcgcctt 258

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	258	100.0	11916	22	AAH79279	Streptomyces averm
2	258	100.0	12381	21	AAZ58381	Streptomyces averm
3	258	100.0	30690	21	AAA92301	S. avermitilis ave
4	258	100.0	30690	22	AAH79277	Streptomyces averm
5	93.4	36.2	125401	22	AAD17186	Streptomyces nours
6	91.8	35.6	43280	18	AAT80413	Tylactone synthase
7	90.2	35.0	53789	19	AAV21187	Amycolatopsis medi
8	85.4	33.1	5676	19	AAV21186	Amycolatopsis medi
9	83.8	32.5	20394	22	AAF24892	Pimaricin biosynth
c 10	83.2	32.2	77536	21	AAA14651	Nucleotide sequenc
11	82.6	32.0	2700	22	AAD17193	ERD48 insert DNA i
12	82.6	32.0	65140	22	AAD17184	Streptomyces nours
13	82.4	31.9	5544	22	AAF24891	Pimaricin biosynth
14	82.2	31.9	44377	18	AAT78508	Platenolide syntha
15	82.2	31.9	44377	18	AAT80414	Platenolide syntha
16	79.4	30.8	11220	21	AAZ87298	S. venezuelae macr
17	79.4	30.8	36778	21	AAZ87318	S. venezuelae pik
18	79.4	30.8	37948	21	AAZ87285	S. venezuelae pik
19	79.4	30.8	38506	21	AAA75633	Nucleotide sequenc
20	79.4	30.8	38506	21	AAZ56001	Recombinant cosmid
21	77.4	30.0	1681	24	AAS18438	Contig 115 DNA enc
22	77.4	30.0	29879	14	AAQ46806	eryA region of S.
23	77.4	30.0	31422	21	AAA92302	S. avermitilis ave
24	77.4	30.0	31422	22	AAH79278	Streptomyces averm
25	74.6	28.9	990	21	AAC55784	Type I polyketide
26	74.6	28.9	12249	21	AAC55840	Complete Mitomycin
27	74.6	28.9	13842	21	AAZ87297	S. venezuelae macr
28	74.6	28.9	18331	21	AAC55857	Complete nucleotid
29	73.6	28.5	4851	22	AAH52062	Mycobacterium tube
c 30	73.6	28.5	4403765	22	AAI99683	Mycobacterium tube
c 31	73.6	28.5	4411529	22	AAI99682	Mycobacterium tube
32	73.4	28.4	16767	22	AAF88339	S. spinosa DNA fra
33	73.4	28.4	25360	22	AAF88314	S. spinosa DNA fra
34	73.4	28.4	29736	22	AAF88317	S. spinosa DNA fra
35	73.4	28.4	80161	20	AAZ21501	DNA fragment of Sa
36	73.2	28.4	15872	21	AAZ87283	S. venezuelae vep
c 37	72.8	28.2	748	22	AAA89129	Polyketide synthas
c 38	72.8	28.2	753	22	AAA89143	Polyketide synthas
c 39	72.6	28.1	1434	24	AAS18434	Contig 132 DNA enc
40	71.6	27.8	47981	22	AAF30757	Micromonospora meg
41	71.4	27.7	27541	22	AAD17185	Streptomyces nours
42	71.2	27.6	15872	18	AAT68715	Streptomyces venez
43	71	27.5	77536	21	AAA14651	Nucleotide sequenc
c 44	70.4	27.3	561	22	AAA89135	Polyketide synthas
45	70.2	27.2	1871	24	ABK11624	S. erythraea monen

ALIGNMENTS

RESULT 1
AAH79279
ID AAH79279 standard; DNA; 11916 BP.
XX
AC AAH79279;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis coding sequence derivative SEQ ID NO: 3.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..11916
FT /*tag= a
FT /product= "AAG65268"
FT /partial
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX

```
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 80161
;   TYPE: DNA
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-1

  Query Match            28.4%;  Score 73.4;  DB 4;  Length 80161;
  Best Local Similarity  65.6%;  Pred. No. 1.5e-08;
  Matches  107;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy         95 GCACCGCGTCGGATGCCGGCCCATCGGCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCA 154
              |||   ||   ||||  | |     |  |  ||| ||| | || || ||||||||  | |
Db      75497 GCAGGTCGAGCGATGTGGACATCGAGAAGCCTTTGGCCGAGCTGGGTTTCGACTCGCTGA 75556

Qy        155 TGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGC 214
               ||   ||||||| ||||||||| |||    |  ||||| || |||||| | |||| |||||
Db      75557 CGGCCATCGAACTCCGCAACCGTCTCGCTACCGCCACCGGACTGCGGCTTCCCGCGACGC 75616

Qy        215 TCATCTTCGACCACACCACGCCGGCCGCGGTCGCCGCGCGCCT 257
              |   ||||||||||  |||  |||| ||  |||||  || | |
Db      75617 TGGCCTTCGACCACGGCACTGCGGCGGCACTCGCCCAGCACGT 75659


RESULT 15
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-1

  Query Match            28.4%;  Score 73.2;  DB 4;  Length 15872;
  Best Local Similarity  64.1%;  Pred. No. 1.5e-08;
  Matches  127;  Conservative    0;  Mismatches   68;  Indels    3;  Gaps    1;

Qy         63 CATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGC 122
              | || |||    |   ||  ||||    || || | |||  || |  || | |   |||
Db       8203 CGTGACGGGTGTCGTCGCGGACGTCCTCGGGCACGCCGACTCCGCCGCGGTCGGGGCGGA 8262

Qy        123 GAG---TTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCT 179
              | |    |||    |||||||||||||||| |    || || || || |||||||| ||
Db       8263 GCGGTCCTTCAAGGACCTCGGCTTCGACTCCCTGGCCGGGGTGGAGCTCCGCAACCGGCT 8322

Qy        180 CAGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGC 239
               | |   ||||| ||||||||| | |||| |||     ||||||||||  || |||||||
Db       8323 GAACGCCGCCACCGGCCTGCGGCTCCCCGCGACCACGGTCTTCGACCATCCCTCGCCGGC 8382

Qy        240 CGCGGTCGCCGCGCGCCT 257
              |||| ||||  | |  ||
Db       8383 CGCGCTCGCGTCCCATCT 8400


Search completed: June 18, 2003, 01:06:20
Job time : 29.1763 secs
```

```
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             28.5%;  Score 73.6;  DB 4;  Length 4403765;
  Best Local Similarity   66.2%;  Pred. No. 1.7e-08;
  Matches  106;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy          93 GCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTTCGCCCACCTCGGCTTCGACTCCGT 152
               || | ||||   ||| | || |||     |    |||| | |||||||| ||||||||  |
Db     3286037 GCCCTCCGCCGAGGACGTCGACCCCGACACCGAATTCGGCGACCTCGGTTTCGACTCATT 3285978

Qy         153 CATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGCGGTTGCCCGTGAC 212
                 | || ||| ||  | ||||||||||||  |  ||||| || |||  | ||||    ||
Db     3285977 AACGGCTGTGGAGTTACGCAACCGCCTCAAAACCGCCACCGGACTGACGCTGCCACCTAC 3285918

Qy         213 GCTCATCTTCGACCACACCACGCCGGCCGCGGTCGCCGCG 252
                 | || ||||| ||  |||| ||  | ||||||||||  |
Db     3285917 CGTGATTTTCGATCATCCCACTCCCACTGCGGTCGCCGAG 3285878


RESULT 12
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             28.5%;  Score 73.6;  DB 4;  Length 4411529;
  Best Local Similarity   66.2%;  Pred. No. 1.7e-08;
  Matches  106;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy          93 GCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTTCGCCCACCTCGGCTTCGACTCCGT 152
               || | ||||   ||| | || |||     |    |||| | ||||||| ||||||||  |
Db     3291761 GCCCTCCGCCGAGGACGTCGACCCCGACACCGAATTCGGCGACCTCGGTTTCGACTCATT 3291702

Qy         153 CATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGCGGTTGCCCGTGAC 212
                 | || ||| ||  | ||||||||||||  |  ||||| || |||  | |.|||    ||
Db     3291701 AACGGCTGTGGAGTTACGCAACCGCCTCAAAACCGCCACCGGACTGACGCTGCCACCTAC 3291642

Qy         213 GCTCATCTTCGACCACACCACGCCGGCCGCGGTCGCCGCG 252
                 | || ||||| ||  |||| ||  | ||||||||||  |
Db     3291641 CGTGATTTTCGATCATCCCACTCCCACTGCGGTCGCCGAG 3291602


RESULT 13
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  80161 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-036-987A-1

  Query Match             28.4%;  Score 73.4;  DB 3;  Length 80161;
  Best Local Similarity   65.6%;  Pred. No. 1.5e-08;
  Matches  107;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy          95 GCACCGCGTCGGATGCCGGCCCATCGGCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCA 154
               |||   ||   ||||  | |     |  |  ||| ||| | || || ||||||||  | |
Db      75497 GCAGGTCGAGCGATGTGGACATCGAGAAGCCTTTGGCCGAGCTGGGTTTCGACTCGCTGA 75556

Qy         155 TGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGC 214
                 ||   ||||||| |||||||| |||   |  ||||| || ||||||  | |||| |||||
Db      75557 CGGCCATCGAACTCCGCAACCGTCTCGCTACCGCCACCGGACTGCGGCTTCCCGCGACGC 75616

Qy         215 TCATCTTCGACCACACCACGCCGGCCGCGGTCGCCGCGCCCT 257
                 |  ||||||||||  |||  |||| ||  |||||  || | |
Db      75617 TGGCCTTCGACCACGGCACTGCGGCGGCACTCGCCCAGCACGT 75659


RESULT 14
US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
```

```
;       LOCATION:  744..6659
;       OTHER INFORMATION:  /function= "APPROXIMATE SPAN OF
;       OTHER INFORMATION:  MODULE 1"
;       OTHER INFORMATION:  /label= FUNCTION
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  744..11219
;       OTHER INFORMATION:  /function= "gene= "eryA""
;       OTHER INFORMATION:  /product= ""ORF1 encoding modules 1 & 2 for
;       OTHER INFORMATION:  6-deoxyerythronolide B""
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  744..1868
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain 1 of module 1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  1998..2198
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain 1 of module 1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  2250..3626
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacylACP synthase domain/module1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3831..4811
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase 2 domain of module 1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  5574..6125
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6369..6626
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain 2 of module 1"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6678..11219
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  module 2"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  6678..8066
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoacyl ACPsynthase of module 2"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  8262..9305
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyltransferase domain of module 2"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  9906..10454
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  beta-ketoreductase domain of module 2"
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  10707..10964
;       OTHER INFORMATION:  /function= "approximate span of
;       OTHER INFORMATION:  acyl carrier domain of module 2"
US-08-439-009A-1

  Query Match            30.0%;  Score 77.4;  DB 3;  Length 11219;
  Best Local Similarity  59.9%;  Pred. No. 1.6e-09;
  Matches  154;  Conservative    0;  Mismatches   91;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||| | |  ||| | |   |  ||| |  | ||  |||  ||  |||||| ||| ||  |

Db      10707 CTGGCGGGTCTGCCGCGTGCCGAGCGGACGGCGGAGCTGGTGCGCCTGGTCCGCACCAGC 10766

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              |||  | | |||||||  | ||||||            | ||  ||  |  || |. |
Db      10767 ACCGCGACCGTGCTGGGCCACGACGA-----------CCCGAAGGCGGTGCGCGCGACC 10814

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
               ||   |||    | ||||| ||||||||  |   ||  |||   ||||||||||  ||| 
Db      10815 ACGCCGTTCAAGGAGCTCGGGTTCGACTCGCTGGCGGCCGTCCGGCTGCGCAACCTGCTC 10874

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              | |  |||||| || || ||  ||||   ||||||  |||||||||||  | |   |  ||
Db      10875 AACGCGGCCACCGGGCTCCGCCTGCCGTCGACGCTGGTCTTCGACCACCCGAACGCCTCC 10934

Qy        241 GCGGTCGCCGCGCGCCT 257
              |||||||||||   |||
Db      10935 GCGGTCGCCGGTTTCCT 10951


RESULT 10
US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-30

  Query Match            28.9%;  Score 74.6;  DB 4;  Length 13842;
  Best Local Similarity  55.6%;  Pred. No. 7.1e-09;
  Matches  143;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||| ||| || |    || |   |  |    ||    |||  |  | | ||  | |
Db       7441 CTGGCCGAGCGGCTGGCCGCCGCGGCTCCCGGCGAGCGTACGGAGATCCTCCTCGGTCTC 7500

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
                 ||||  |      || ||           |||  |  ||  ||| | || | |
Db       7501 GTACGGGCGCAGGCCGCCGCCGTGCTCCGGATGCGTTCGCCGGAGGACGTCGCCGCCGAC 7560

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                |||    || |||||||||||||  ||   |||||||| ||||||||| | ||
Db       7561 CGCGCCTTCAAGGACATCGGCTTCGACTCGCTCGCCGGTGTCGAGCTGCGCAACAGGCTG 7620

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              | |  ||| || || || | | |||||| ||||||| |||||||||| | ||||||
Db       7621 ACCCGGGCGACCGGGCTCCAGCTGCCCGCGACGCTCGTCTTCGACCACCCGACGCCGCTG 7680

Qy        241 GCGGTCGCCGCGCGCCT 257
              ||  |||   |||  ||
Db       7681 GCCCTCGTGTCGCTGCT 7697


RESULT 11
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
```

```
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 2338
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..6659
;      OTHER INFORMATION:  /function= "APPROXIMATE SPAN OF
;      OTHER INFORMATION:  MODULE 1"
;      OTHER INFORMATION:  /label= FUNCTION
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  744..11219
;      OTHER INFORMATION:  /function= "gene= "eryA""
;      OTHER INFORMATION:  /product= ""ORF1 encoding modules 1 & 2 for
;      OTHER INFORMATION:  6-deoxyerythronolide B""
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  744..1868
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1998..2198
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 1 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  2250..3626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacylACP synthase domain/module1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  3831..4811
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase 2 domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  5574..6125
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6369..6626
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain 2 of module 1"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..11219
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  6678..8066
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoacyl ACPsynthase of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  8262..9305
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyltransferase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  9906..10454
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  beta-ketoreductase domain of module 2"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  10707..10964
;      OTHER INFORMATION:  /function= "approximate span of
;      OTHER INFORMATION:  acyl carrier domain of module 2"
US-07-642-734C-1
```

```
  Query Match           30.0%;  Score 77.4;  DB 1;  Length 11219;
  Best Local Similarity   59.9%;  Pred. No. 1.6e-09;
  Matches  154;  Conservative    0;  Mismatches   91;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||| | |  ||| | |    |  ||| |  | ||  |||  ||  |||||| ||| ||  |
Db      10707 CTGGCGGGTCTGCCGCGTGCCGAGCGGACGGCGGAGCTGGTGCGCCTGGTCCGCACCAGC 10766

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
              |||  | | ||||||  | ||||||                | ||  ||  |   || | |
Db      10767 ACCGCGACCGTGCTGGGCCACGACGA-----------CCCGAAGGCGGTGCGCGCGACC 10814

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              ||   |||    | ||||| ||||||||  |   ||  |||   ||||||||||  |||
Db      10815 ACGCCGTTCAAGGAGCTCGGGTTCGACTCGCTGGCGGCCGTCCGGCTGCGCAACCTGCTC 10874

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              | |  |||||| || || ||  ||||   ||||||  |||||||||||| | |   |  ||
Db      10875 AACGCGGCCACCGGGCTCCGCCTGCCGTCGACGCTGGTCTTCGACCACCCGAACGCCTCC 10934

Qy        241 GCGGTCGCCGCGCGCCT 257
              |||||||||||   |||
Db      10935 GCGGTCGCCGGTTTCCT 10951
```

```
RESULT 9
US-08-439-009A-1
; Sequence 1, Application US/08439009A
; Patent No. 6004787
;  GENERAL INFORMATION:
;    APPLICANT:  Donadio, S
;    APPLICANT:  Katz, L
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Method of Directing Biosynthesis of
;    TITLE OF INVENTION:  Specific Polyketides
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steven F. Weinstock
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,009A
;      FILING DATE:  11-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Casuto, Dianne
;      REGISTRATION NUMBER:  40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847-938-3137
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11219 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharopolyspora erythraea
;      STRAIN:  NRRL 2338
;    FEATURE:
;      NAME/KEY:  misc_feature
```

```
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-19

  Query Match             30.8%;  Score 79.4;  DB 3;  Length 38506;
  Best Local Similarity   60.2%;  Pred. No. 6.2e-10;
  Matches  150;  Conservative    0;  Mismatches   96;  Indels    3;  Gaps    1;

Qy          9 GCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGC 68
              |||  || | | |   | | |    | | ||  | |||||  |  |  |  |   | |
Db      24506 GCTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCA 24565

Qy         69 GGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTT 128
               ||| || || |   ||| |||   ||||| ||       |  | ||    |  |  ||
Db      24566 CGTGGCGACCGTCCTGGGACACG---GCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTT 24622

Qy        129 CGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGC 188
              |  | ||  ||| ||||||||  |||  |  ||||||| |||||||| |||| |   ||
Db      24623 CCGCGACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGC 24682

Qy        189 CACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGC 248
               || || |||||| |||| |  |||||  |||||||||| |||| ||||  | | ||||
Db      24683 GACCGGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGC 24742

Qy        249 CGCGCGCCT 257
              || || |||
Db      24743 CGGGCACCT 24751


RESULT 7
US-09-434-288-7
; Sequence 7, Application US/09434288
; Patent No. 6303767
; GENERAL INFORMATION:
;  APPLICANT:  Betlach C., Melanie
;  APPLICANT:  McDaniel, Robert
;  TITLE OF INVENTION:  POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
;  TITLE OF INVENTION:  CONSTRUCTS THEREFOR
;  FILE REFERENCE:  30062-20030.00
;  CURRENT APPLICATION NUMBER: US/09/434,288
;  CURRENT FILING DATE:  1999-11-05
;  PRIOR APPLICATION NUMBER:  60/107,093
;  PRIOR FILING DATE:  1998-11-05
;  NUMBER OF SEQ ID NOS:  12
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 7
;   LENGTH: 1681
;   TYPE: DNA
;   ORGANISM: Streptomyces narbonensis
US-09-434-288-7

  Query Match             30.0%;  Score 77.4;  DB 4;  Length 1681;
  Best Local Similarity   59.9%;  Pred. No. 1.5e-09;
  Matches  154;  Conservative    0;  Mismatches   91;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||  ||||| || ||||      |   |     |||  || | |||||| || ||  | |||
Db        541 CTCGCCGCGATGACCCCGGACGACAGGACCGCGCACCTGCGGGACCTCGTCCGTACGCAC 600

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
                | | ||  |||    |||              || || ||       | || ||    |
Db        601 GTGGCGACCGTCCTGGGACACG-----------GCGCCCCGAGCCGGGTCGACCTGGAG 648

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                  |||  | ||  ||| |||||||||  |||  |  ||||| || ||||||||||||
Db        649 CGCGCCTTCCGCGACACCGGTTTCGACTCCCTCACCGCCGTCGAGCTCCGCAACCGCCTC 708

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              | |   ||||| || |||||  | || |  |||||| |||||||||| |||| ||||
Db        709 AACGCCGCCACCGGGCTGCGCCTCCCGGCCACGCTCGTCTTCGACCACCCCACTCCGGGG 768

Qy        241 GCGGTCGCCGCGCGCCT 257
              | | |||||| || |||
Db        769 GAGCTCGCCGGGCACCT 785


RESULT 8
US-07-642-734C-1
; Sequence 1, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
;   APPLICANT:  Katz, L
;   APPLICANT:  Donadio, S
;   APPLICANT:  Mcalpine, J B
;   TITLE OF INVENTION:  Recombinant DNA Method for Producing
;   TITLE OF INVENTION:  Erythromycin Analogs
;   NUMBER OF SEQUENCES:  27
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Edward H. Gorman
;     STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;     STREET:  Park Rd
;     CITY:  Abbott Park
;     STATE:  IL
;     COUNTRY:  US
;     ZIP:  60064-3500
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/07/642,734C
;     FILING DATE:  17-JAN-91
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Danckers, Andreas M
;     REGISTRATION NUMBER:  32652
;     REFERENCE/DOCKET NUMBER:  4952.US.01
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  708-937-9396
;     TELEFAX:  708-938-2623
; INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  11219 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  unknown
;   MOLECULE TYPE:  DNA (genomic)
;   HYPOTHETICAL:  NO
```

```
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-198-1

  Query Match             31.9%;  Score 82.2;  DB 2;  Length 44377;
  Best Local Similarity   61.1%;  Pred. No. 1.5e-10;
  Matches  157;  Conservative    0;  Mismatches   88;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||  | |  |  ||| |  |  |||    | | || || || ||||||||||  || |
Db      19476 CTCGCGGGACGTTCCGCACCCGAGCGTACGGCTCTCCTGCTCGACCTGGTGCGGACCGAG 19535

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
                |  |||||||||      |||             ||  | |  | |    |||| |  |
Db      19536 GTCGCGGCGGTGCTCGGACACG-----------GCGACCCCGCCGCGATCGGCGCCGCC 19583

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                |  |||    ||  ||| |||||||||| |||  | ||||||  || ||||||||  ||
Db      19584 CGCACCTTCAAGGACGCCGGATTCGACTCCCTCACCGCTGTCGACCTCCGCAACCGGCTG 19643

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              | ||    ||| || |||||| |||||| ||| ||| |||||||||| |||| |||  |
Db      19644 AACACACGCACCGGACTGCGGCTGCCCGCGACCCTCGTCTTCGACCACCCCACACCGCTC 19703

Qy        241 GCGGTCGCCGCGCGCCT 257
              ||  ||||||  | |||
Db      19704 GCCCTCGCCGAACTCCT 19720


RESULT 4
US-09-105-537-32
; Sequence 32, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-32

  Query Match             30.8%;  Score 79.4;  DB 4;  Length 11220;
  Best Local Similarity   60.2%;  Pred. No. 5.8e-10;
  Matches  150;  Conservative    0;  Mismatches   96;  Indels    3;  Gaps    1;

Qy          9 GCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGC 68
              |||  || | | |   | | |    | | ||  | |||||  |  |  |  |   | |
Db      10677 GCTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCA 10736

Qy         69 GGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTT 128
               |||  || || |   ||| |||   ||||| ||       |  | ||    |  |   ||
Db      10737 CGTGGCGACCGTCCTGGGACACG---GCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTT 10793

Qy        129 CGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGC 188
              |  | ||  ||| ||||||||  |||  |  ||||||| ||||||||  |||| |   ||
Db      10794 CCGCGACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGC 10853

Qy        189 CACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGC 248
              || || |||||| |||| |  |||||  |||||||||| |||| ||||  | | ||||
Db      10854 GACCGGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGC 10913

Qy        249 CGCGCGCCT 257
              || || |||
Db      10914 CGGGCACCT 10922


RESULT 5
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-5

  Query Match             30.8%;  Score 79.4;  DB 4;  Length 36778;
  Best Local Similarity   60.2%;  Pred. No. 6.2e-10;
  Matches  150;  Conservative    0;  Mismatches   96;  Indels    3;  Gaps    1;

Qy          9 GCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGC 68
              |||  || | | |   | | |    | | ||  | |||||  |  |  |  |   | |
Db      26364 GCTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCA 26423

Qy         69 GGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTT 128
               |||  || || |   ||| |||   ||||| ||       |  | ||    |  |   ||
Db      26424 CGTGGCGACCGTCCTGGGACACG---GCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTT 26480

Qy        129 CGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGC 188
              |  | ||  ||| ||||||||  |||  |  ||||||| ||||||||  |||| |   ||
Db      26481 CCGCGACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGC 26540

Qy        189 CACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGC 248
              || || |||||| |||| |  |||||  |||||||||| |||| ||||  | | ||||
Db      26541 GACCGGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGC 26600

Qy        249 CGCGCGCCT 257
              || || |||
Db      26601 CGGGCACCT 26609


RESULT 6
US-09-320-878-19
; Sequence 19, Application US/09320878A
```

```
;      NAME/KEY:  CDS
;      LOCATION:  31232..36067
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36249..41774
US-08-804-227C-1

  Query Match             35.6%;  Score 91.8;  DB 2;  Length 43280;
  Best Local Similarity   63.4%;  Pred. No. 9.8e-13;
  Matches  163;  Conservative    0;  Mismatches   82;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||||||| |||  |  |  | |   |||  | ||  |||||| |  | | |||||||||
Db      30648 CTGACCGGGCTCACGGCACAGGAAGGGCACGCCCTCGTCCTGGCCGAGATCCGCGCCCAC 30707

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
               |  ||||||||||  | ||| |    ||||    | | |   |||    | |
Db      30708 GCGGCGGCGGTGCTGGGCCACGGCTCCGACGACTCGATCCCCGAGGACCGGGCC------ 30761

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                 |||    |||||||||||||||||  |||  |  || ||  |||||||||| ||
Db      30762 ------TTCAAGGACCTCGGCTTCGACTCGCTCACCGCCGTGGAGATGCGCAACCGGCTG 30815

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              |||  |||||| ||||| ||| ||||||  || ||| |||||||||| | || |||| |
Db      30816 AGCGCGGCCACCGGCCTCCGGCTGCCCGCCACCCTCGTCTTCGACCACCCGACCCCGGGC 30875

Qy        241 GCGGTCGCCGCGCGCCT 257
              | | | ||||  | |||
Db      30876 GAGCTGGCCGGCCACCT 30892


RESULT 2
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-227C-7

  Query Match             31.9%;  Score 82.2;  DB 2;  Length 44377;
  Best Local Similarity   61.1%;  Pred. No. 1.5e-10;
  Matches  157;  Conservative    0;  Mismatches   88;  Indels   12;  Gaps    1;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||  | |  |  ||| |  |  |||    | | || || || ||||||||||  || |
Db      19476 CTCGCGGGACGTTCCGCACCCGAGCGTACGGCTCTCCTGCTCGACCTGGTGCGGACCGAG 19535

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
               |  |||||||||      |||              ||  | |  | |    |||| |  |
Db      19536 GTCGCGGCGGTGCTCGGACACG------------GCGACCCCGCCGCGATCGGCGCCGCC 19583

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
                 |  |||     ||  ||| ||||||||| |||   | |||||  || ||||||||  ||
Db      19584 CGCACCTTCAAGGACGCCGGATTCGACTCCCTCACCGCTGTCGACCTCCGCAACCGGCTG 19643

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              | ||     ||| || |||||| |||||| ||| ||| ||||||||||| |||| |||  |
Db      19644 AACACACGCACCGGACTGCGGCTGCCCGCGACCCTCGTCTTCGACCACCCCACACCGCTC 19703

Qy        241 GCGGTCGCCGCGCGCCT 257
              ||  ||||||  | |||
Db      19704 GCCCTCGCCGAACTCCT 19720


RESULT 3
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 11.1763 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_1096_1353
Perfect score: 258
Sequence: 1 ctgaccgcgctgtcccccac.....ccgcggtcgccgcgcgcctt 258

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	91.8	35.6	43280	2	US-08-804-227C-1	Sequence 1, Appli
2	82.2	31.9	44377	2	US-08-804-227C-7	Sequence 7, Appli
3	82.2	31.9	44377	2	US-08-804-198-1	Sequence 1, Appli
4	79.4	30.8	11220	4	US-09-105-537-32	Sequence 32, Appl
5	79.4	30.8	36778	4	US-09-105-537-5	Sequence 5, Appli
6	79.4	30.8	38506	3	US-09-320-878-19	Sequence 19, Appl
7	77.4	30.0	1681	4	US-09-434-288-7	Sequence 7, Appli
8	77.4	30.0	11219	1	US-07-642-734C-1	Sequence 1, Appli
9	77.4	30.0	11219	3	US-08-439-009A-1	Sequence 1, Appli
10	74.6	28.9	13842	4	US-09-105-537-30	Sequence 30, Appl
c 11	73.6	28.5	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c 12	73.6	28.5	4411529	4	US-09-103-840A-1	Sequence 1, Appli
13	73.4	28.4	80161	3	US-09-036-987A-1	Sequence 1, Appli
14	73.4	28.4	80161	4	US-09-370-700-1	Sequence 1, Appli
15	73.2	28.4	15872	4	US-09-105-537-1	Sequence 1, Appli
c 16	72.8	28.2	748	3	US-09-154-083-1	Sequence 1, Appli
c 17	72.8	28.2	753	3	US-09-154-083-28	Sequence 28, Appl
c 18	72.6	28.1	1434	4	US-09-434-288-3	Sequence 3, Appli
19	70.6	27.4	20235	1	US-07-642-734C-3	Sequence 3, Appli
20	70.6	27.4	20235	3	US-08-439-009A-3	Sequence 3, Appli
c 21	70.4	27.3	561	3	US-09-154-083-13	Sequence 13, Appl
22	69	26.7	50937	4	US-09-428-517-1	Sequence 1, Appli
23	68.6	26.6	4403765	4	US-09-103-840A-2	Sequence 2, Appli
24	68.6	26.6	4411529	4	US-09-103-840A-1	Sequence 1, Appli
25	68.4	26.5	4689	4	US-09-105-537-34	Sequence 34, Appl
26	67.6	26.2	13987	2	US-08-804-227C-13	Sequence 13, Appl
27	64.6	25.0	33529	4	US-09-144-085-3	Sequence 3, Appli
28	63.2	24.5	28958	1	US-08-258-261B-6	Sequence 6, Appli
29	63.2	24.5	28958	1	US-08-456-837-6	Sequence 6, Appli
30	63.2	24.5	28958	1	US-08-457-342-6	Sequence 6, Appli
31	63.2	24.5	28958	1	US-08-457-646A-6	Sequence 6, Appli
32	63.2	24.5	28958	1	US-08-458-076A-6	Sequence 6, Appli
33	63.2	24.5	28958	1	US-08-764-233A-4	Sequence 4, Appli
34	63.2	24.5	28958	1	US-08-457-335A-6	Sequence 6, Appli
35	63.2	24.5	28958	1	US-08-729-214-6	Sequence 6, Appli
36	63.2	24.5	28958	3	US-09-028-934-6	Sequence 6, Appli
37	63.2	24.5	49377	1	US-08-764-233A-1	Sequence 1, Appli
38	58.6	22.7	4041	4	US-09-105-537-36	Sequence 36, Appl
39	55	21.3	68750	3	US-09-335-409-1	Sequence 1, Appli
40	55	21.3	68750	4	US-09-568-102-1	Sequence 1, Appli
41	55	21.3	68750	4	US-09-567-969-1	Sequence 1, Appli
42	55	21.3	68750	4	US-09-568-480-1	Sequence 1, Appli
43	55	21.3	68750	4	US-09-568-486-1	Sequence 1, Appli
44	55	21.3	68750	4	US-09-568-472-1	Sequence 1, Appli
45	55	21.3	68750	4	US-09-567-899-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rosteck, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCI(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..14234
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14351..19945
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20010..31199
; FEATURE:

```
                   |||    || ||||||||||||||  ||   |||||||| ||||||||| | ||
Db       7561 CGCGCCTTCAAGGACATCGGCTTCGACTCGCTCGCCGGTGTCGAGCTGCGCAACAGGCTG 7620

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              | |  ||| || || || | | |||||| ||||||| |||||||||| | ||||||
Db       7621 ACCCGGGCGACCGGGCTCCAGCTGCCCGCGACGCTCGTCTTCGACCACCCGACGCCGCTG 7680

Qy        241 GCGGTCGCCGCGCGCCT 257
              ||  |||   ||| .||
Db       7681 GCCCTCGTGTCGCTGCT 7697


RESULT 14
US-09-712-363-116
; Sequence 116, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
;  APPLICANT: Eisenberg, David
;  APPLICANT:  Rotstein, Sergio H.
;  APPLICANT:  Marcotte, Edward M.
;  TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
;  TITLE OF INVENTION:  INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
;  FILE REFERENCE: 07419-032001
;  CURRENT APPLICATION NUMBER: US/09/712,363
;  CURRENT FILING DATE:  2000-11-13
;  PRIOR APPLICATION NUMBER: PCT/US00/02246
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: 60/179,531
;  PRIOR FILING DATE: 2000-02-01
;  PRIOR APPLICATION NUMBER: 60/117,844
;  PRIOR FILING DATE: 1999-01-29
;  PRIOR APPLICATION NUMBER: 60/118,206,
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: 60/126,593
;  PRIOR FILING DATE: 1999-03-26
;  PRIOR APPLICATION NUMBER: 60/134,093
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: 60/134,092
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: 60/165,124
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/165,086
;  PRIOR FILING DATE: 1999-11-12
;  NUMBER OF SEQ ID NOS: 292
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 116
;   LENGTH: 4851
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
US-09-712-363-116

  Query Match             28.5%;  Score 73.6;  DB 9;  Length 4851;
  Best Local Similarity   66.2%;  Pred. No. 6.8e-12;
  Matches  106;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy         93 GCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTTCGCCCACCTCGGCTTCGACTCCGT 152
              || | ||||   ||| | || |||     |    |||| | ||||||| ||||||||  |
Db       4590 GCCCTCCGCCGAGGACGTCGACCCCGACACCGAATTCGGCGACCTCGGTTTCGACTCATT 4649

Qy        153 CATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGCGGTTGCCCGTGAC 212
               | || ||| ||  | ||||||||||||||  |  ||||| || |||  | ||||    ||
Db       4650 AACGGCTGTGGAGTTACGCAACCGCCTCAAAACCGCCACCGGACTGACGCTGCCACCTAC 4709

Qy        213 GCTCATCTTCGACCACACCACGCCGGCCGCGGTCGCCGCG 252
              | || ||||| ||  |||| ||  | ||||||||||| |
Db       4710 CGTGATTTTCGATCATCCCACTCCCACTGCGGTCGCCGAG 4749


RESULT 15
US-09-860-846-1
; Sequence 1, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-1

  Query Match             28.4%;  Score 73.2;  DB 9;  Length 15872;
  Best Local Similarity   64.1%;  Pred. No. 8.6e-12;
  Matches  127;  Conservative    0;  Mismatches   68;  Indels    3;  Gaps    1;

Qy         63 CATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGC 122
              | || |||    |   ||  ||||    || || | |||  || |  || | |    |||
Db       8203 CGTGACGGGTGTCGTCGCGGACGTCCTCGGGCACGCCGACTCCGCCGCGGTCGGGGCGGA 8262

Qy        123 GAG---TTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCT 179
              | |    |||    |||||||||||||||||| |    || || || || |||||||| ||
Db       8263 GCGGTCCTTCAAGGACCTCGGCTTCGACTCCCTGGCCGGGGTGGAGCTCCGCAACCGGCT 8322

Qy        180 CAGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGC 239
              | |   ||||| ||||||||| | |||| |||      ||||||||||  || |||||||
Db       8323 GAACGCCGCCACCGGCCTGCGGCTCCCCGCGACCACGGTCTTCGACCATCCCTCGCCGGC 8382

Qy        240 CGCGGTCGCCGCGCGCCT 257
              |||| ||||  | |  ||
Db       8383 CGCGCTCGCGTCCCATCT 8400


Search completed: June 18, 2003, 05:08:05
Job time : 31.5809 secs
```

```
Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              |||     || |||||||||||||  ||   |||||||| |||||||| | ||
Db       7561 CGCGCCTTCAAGGACATCGGCTTCGACTCGCTCGCCGGTGTCGAGCTGCGCAACAGGCTG 7620

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              | |  ||| || || || | | |||||| ||||||| |||||||||| | ||||||
Db       7621 ACCCGGGCGACCGGGCTCCAGCTGCCCGCGACGCTCGTCTTCGACCACCCGACGCCGCTG 7680

Qy        241 GCGGTCGCCGCGCGCCT 257
              ||  |||   |||  ||
Db       7681 GCCCTCGTGTCGCTGCT 7697


RESULT 11
US-09-988-384B-30
; Sequence 30, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-30

  Query Match             28.9%;  Score 74.6;  DB 9;  Length 13842;
  Best Local Similarity   55.6%;  Pred. No. 3.5e-12;
  Matches  143;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||| ||| || |    || |   |   |        ||    |||  |  | | ||  | |
Db       7441 CTGGCCGAGCGGCTGGCCGCCGCGGCTCCCGGCGAGCGTACGGAGATCCTCCTCGGTCTC 7500

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
                ||||  |.    || ||          |||  |  ||  ||| | || | | |
Db       7501 GTACGGGCGCAGGCCGCCGCCGTGCTCCGGATGCGTTCGCCGGAGGACGTCGCCGCCGAC 7560

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              |||     || |||||||||||||  ||   |||||||| |||||||| | ||
Db       7561 CGCGCCTTCAAGGACATCGGCTTCGACTCGCTCGCCGGTGTCGAGCTGCGCAACAGGCTG 7620

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              | |  ||| || || || | | |||||| ||||||| |||||||||| | ||||||
Db       7621 ACCCGGGCGACCGGGCTCCAGCTGCCCGCGACGCTCGTCTTCGACCACCCGACGCCGCTG 7680

Qy        241 GCGGTCGCCGCGCGCCT 257
              ||  |||   |||  ||
Db       7681 GCCCTCGTGTCGCTGCT 7697


RESULT 12
US-09-836-821-30
; Sequence 30, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-30

  Query Match             28.9%;  Score 74.6;  DB 9;  Length 13842;
  Best Local Similarity   55.6%;  Pred. No. 3.5e-12;
  Matches  143;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||| ||| || |    || |   |   |        ||    |||  |  | | ||  | |
Db       7441 CTGGCCGAGCGGCTGGCCGCCGCGGCTCCCGGCGAGCGTACGGAGATCCTCCTCGGTCTC 7500

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
                ||||  |    || ||          |||  |  ||  ||| | || | | |
Db       7501 GTACGGGCGCAGGCCGCCGCCGTGCTCCGGATGCGTTCGCCGGAGGACGTCGCCGCCGAC 7560

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
              |||     || |||||||||||||  ||   |||||||| |||||||||| | ||
Db       7561 CGCGCCTTCAAGGACATCGGCTTCGACTCGCTCGCCGGTGTCGAGCTGCGCAACAGGCTG 7620

Qy        181 AGCAAGGCCACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCC 240
              | |  ||| || || || | | |||||| ||||||| |||||||||| | ||||||
Db       7621 ACCCGGGCGACCGGGCTCCAGCTGCCCGCGACGCTCGTCTTCGACCACCCGACGCCGCTG 7680

Qy        241 GCGGTCGCCGCGCGCCT 257
              ||  |||   |||  ||
Db       7681 GCCCTCGTGTCGCTGCT 7697


RESULT 13
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-30

  Query Match             28.9%;  Score 74.6;  DB 10;  Length 13842;
  Best Local Similarity   55.6%;  Pred. No. 3.5e-12;
  Matches  143;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||| ||| || |    || |   |   |        ||    |||  |  | | ||  | |
Db       7441 CTGGCCGAGCGGCTGGCCGCCGCGGCTCCCGGCGAGCGTACGGAGATCCTCCTCGGTCTC 7500

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
                ||||  |    || ||          |||  |  ||  ||| | || | | |
Db       7501 GTACGGGCGCAGGCCGCCGCCGTGCTCCGGATGCGTTCGCCGGAGGACGTCGCCGCCGAC 7560

Qy        121 GCGAGTTTCGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTC 180
```

```
Db      26601 CGGGCACCT 26609


RESULT 8
US-09-988-384B-5
; Sequence 5, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 5
;   LENGTH: 37948
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-5

  Query Match             30.8%;  Score 79.4;  DB 9;  Length 37948;
  Best Local Similarity   60.2%;  Pred. No. 1.5e-13;
  Matches  150;  Conservative    0;  Mismatches   96;  Indels    3;  Gaps    1;

Qy          9 GCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGC 68
              |||  || | | |   | | |    | | ||  | |||||  |  |  |  |  | |
Db      27534 GCTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCA 27593

Qy         69 GGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTT 128
               |||  || || |   ||| |||   ||||| ||       |  | ||     |  |    ||
Db      27594 CGTGGCGACCGTCCTGGGACACG---GCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTT 27650

Qy        129 CGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGC 188
              |  | ||  ||| |||||||||  |||  |  ||||||| ||||||| |||| |    ||
Db      27651 CCGCGACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGC 27710

Qy        189 CACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGC 248
               || || |||||| |||| |  |||||  |||||||||| |||| ||||   | | ||||
Db      27711 GACCGGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGC 27770

Qy        249 CGCGCGCCT 257
              || || |||
Db      27771 CGGGCACCT 27779


RESULT 9
US-09-793-708-19
; Sequence 19, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-19

  Query Match             30.8%;  Score 79.4;  DB 9;  Length 38506;
  Best Local Similarity   60.2%;  Pred. No. 1.5e-13;
  Matches  150;  Conservative    0;  Mismatches   96;  Indels    3;  Gaps    1;

Qy          9 GCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGC 68
              |||  || | | |   | | |    | | ||  | |||||  |  |  |  |  | |
Db      24506 GCTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCA 24565

Qy         69 GGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTT 128
               |||  || || |   ||| |||   ||||| ||       |  | ||     |  |    ||
Db      24566 CGTGGCGACCGTCCTGGGACACG---GCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTT 24622

Qy        129 CGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGC 188
              |  | ||  ||| |||||||||  |||  |  ||||||| ||||||| |||| |    ||
Db      24623 CCGCGACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGC 24682

Qy        189 CACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGC 248
               || || |||||| |||| |  |||||  |||||||||| |||| ||||   | | ||||
Db      24683 GACCGGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGC 24742

Qy        249 CGCGCGCCT 257
              || || |||
Db      24743 CGGGCACCT 24751


RESULT 10
US-09-860-846-30
; Sequence 30, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-30

  Query Match             28.9%;  Score 74.6;  DB 9;  Length 13842;
  Best Local Similarity   55.6%;  Pred. No. 3.5e-12;
  Matches  143;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy          1 CTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCAC 60
              ||| ||| || |    || |   |   |        ||    |||  |  | | ||  | |
Db       7441 CTGGCCGAGCGGCTGGCCGCCGCGGCTCCCGGCGAGCGTACGGAGATCCTCCTCGGTCTC 7500

Qy         61 ACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCG 120
                  ||||  |  . || ||           |||  |  ||  ||| | || | |
Db       7501 GTACGGGCGCAGGCCGCCGCCGTGCTCCGGATGCGTTCGCCGGAGGACGTCGCCGCCGAC 7560
```

```
Qy        249 CGCGCGCCT 257
              || || |||
Db      10914 CGGGCACCT 10922
```

RESULT 5
US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-5

Query Match 30.8%; Score 79.4; DB 9; Length 36778;
Best Local Similarity 60.2%; Pred. No. 1.5e-13;
Matches 150; Conservative 0; Mismatches 96; Indels 3; Gaps 1;

```
Qy          9 GCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGC 68
              ||| || | | |   | | |    | | ||  | |||||  |  |  |  |   | |
Db      26364 GCTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCA 26423

Qy         69 GGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTT 128
               ||| || || |   ||| |||   |||||  ||       |  | ||    |  |   ||
Db      26424 CGTGGCGACCGTCCTGGGACACG---GCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTT 26480

Qy        129 CGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGC 188
              |  | ||  ||| ||||||||  |||  |  ||||||||  ||||||||  ||||  |   ||
Db      26481 CCGCGACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGC 26540

Qy        189 CACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGC 248
              || || |||||| |||| |  |||||  ||||||||||| |||| ||||  | | ||||
Db      26541 GACCGGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGC 26600

Qy        249 CGCGCGCCT 257
              || || |||
Db      26601 CGGGCACCT 26609

RESULT 6
US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-5

Query Match 30.8%; Score 79.4; DB 9; Length 36778;
Best Local Similarity 60.2%; Pred. No. 1.5e-13;
Matches 150; Conservative 0; Mismatches 96; Indels 3; Gaps 1;

```
Qy 9 GCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGC 68
 ||| || | | | | | | | | || | ||||| | | | | | |
Db 26364 GCTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCA 26423

Qy 69 GGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTT 128
 ||| || || | ||| ||| ||||| || | | || | | ||
Db 26424 CGTGGCGACCGTCCTGGGACACG---GCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTT 26480

Qy 129 CGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGC 188
 | | || ||| |||||||| ||| | |||||||| |||||||| |||| | ||
Db 26481 CCGCGACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGC 26540

Qy 189 CACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGC 248
 || || |||||| |||| | ||||| ||||||||||| |||| |||| | | ||||
Db 26541 GACCGGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGC 26600

Qy 249 CGCGCGCCT 257
 || || |||
Db 26601 CGGGCACCT 26609

RESULT 7
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5

Query Match 30.8%; Score 79.4; DB 10; Length 36778;
Best Local Similarity 60.2%; Pred. No. 1.5e-13;
Matches 150; Conservative 0; Mismatches 96; Indels 3; Gaps 1;

```
Qy          9 GCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGC 68
              ||| || | | |   | | |    | | ||  | |||||  |  |  |  |   | |
Db      26364 GCTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCA 26423

Qy         69 GGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTT 128
               ||| || || |   ||| |||   |||||  ||       |  | ||    |  |   ||
Db      26424 CGTGGCGACCGTCCTGGGACACG---GCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTT 26480

Qy        129 CGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGC 188
              |  | ||  ||| ||||||||  |||  |  ||||||||  ||||||||  ||||  |   ||
Db      26481 CCGCGACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGC 26540

Qy        189 CACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGC 248
              || || |||||| |||| |  |||||  ||||||||||| |||| ||||  | | ||||
Db      26541 GACCGGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGC 26600

Qy        249 CGCGCGCCT 257
              || || |||
```

```
Qy        249 CGCGCGCCT 257
              || || |||
Db      10914 CGGGCACCT 10922


RESULT 2
US-09-988-384B-32
; Sequence 32, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-32

  Query Match            30.8%;  Score 79.4;  DB 9;  Length 11220;
  Best Local Similarity  60.2%;  Pred. No. 1.6e-13;
  Matches  150;  Conservative    0;  Mismatches   96;  Indels    3;  Gaps    1;

Qy          9 GCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGC 68
              ||| || | | |   | | |     | | ||  | |||||  |  |  |  |   | |
Db      10677 GCTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCA 10736

Qy         69 GGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTT 128
               |||  || || |   ||| |||   ||||| ||      |  | ||    |  |   ||
Db      10737 CGTGGCGACCGTCCTGGGACACG---GCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTT 10793

Qy        129 CGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGC 188
              |  | ||  ||| |||||||||  |||  |  |||||||| |||||||| |||| |   ||
Db      10794 CCGCGACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGC 10853

Qy        189 CACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGC 248
              || || |||||| |||| |  |||||  |||||||||| |||| ||||  | | ||||
Db      10854 GACCGGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGC 10913

Qy        249 CGCGCGCCT 257
              || || |||
Db      10914 CGGGCACCT 10922


RESULT 3
US-09-836-821-32
; Sequence 32, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-32

  Query Match            30.8%;  Score 79.4;  DB 9;  Length 11220;
  Best Local Similarity  60.2%;  Pred. No. 1.6e-13;
  Matches  150;  Conservative    0;  Mismatches   96;  Indels    3;  Gaps    1;

Qy          9 GCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGC 68
              ||| || | | |   | | |     | | ||  | |||||  |  |  |  |   | |
Db      10677 GCTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCA 10736

Qy         69 GGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTT 128
               |||  || || |   ||| |||   ||||| ||      |  | ||    |  |   ||
Db      10737 CGTGGCGACCGTCCTGGGACACG---GCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTT 10793

Qy        129 CGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGC 188
              |  | ||  ||| |||||||||  |||  |  |||||||| |||||||| |||| |   ||
Db      10794 CCGCGACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGC 10853

Qy        189 CACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGC 248
              || || |||||| |||| |  |||||  |||||||||| |||| ||||  | | ||||
Db      10854 GACCGGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGC 10913

Qy        249 CGCGCGCCT 257
              || || |||
Db      10914 CGGGCACCT 10922


RESULT 4
US-09-861-289-32
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-32

  Query Match            30.8%;  Score 79.4;  DB 10;  Length 11220;
  Best Local Similarity  60.2%;  Pred. No. 1.6e-13;
  Matches  150;  Conservative    0;  Mismatches   96;  Indels    3;  Gaps    1;

Qy          9 GCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGC 68
              ||| || | | |   | | |     | | ||  | |||||  |  |  |  |   | |
Db      10677 GCTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCA 10736

Qy         69 GGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTT 128
               |||  || || |   ||| |||   ||||| ||      |  | ||    |  |   ||
Db      10737 CGTGGCGACCGTCCTGGGACACG---GCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTT 10793

Qy        129 CGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGC 188
              |  | ||  ||| |||||||||  |||  |  |||||||| |||||||| |||| |   ||
Db      10794 CCGCGACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGC 10853

Qy        189 CACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGC 248
              || || |||||| |||| |  |||||  |||||||||| |||| ||||  | | ||||
Db      10854 GACCGGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGC 10913
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 17:56:34 ; Search time 30.49 Seconds
(without alignments)
12253.192 Million cell updates/sec

Title: US-09-914-286-1_COPY_1096_1353
Perfect score: 258
Sequence: 1 ctgaccgcgctgtcccccac.....ccgcggtcgccgcgcgcctt 258

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	79.4	30.8	11220	9	US-09-860-846-32	Sequence 32, Appl
2	79.4	30.8	11220	9	US-09-988-384B-32	Sequence 32, Appl
3	79.4	30.8	11220	9	US-09-836-821-32	Sequence 32, Appl
4	79.4	30.8	11220	10	US-09-861-289-32	Sequence 32, Appl
5	79.4	30.8	36778	9	US-09-860-846-5	Sequence 5, Appli
6	79.4	30.8	36778	9	US-09-836-821-5	Sequence 5, Appli
7	79.4	30.8	36778	10	US-09-861-289-5	Sequence 5, Appli
8	79.4	30.8	37948	9	US-09-988-384B-5	Sequence 5, Appli
9	79.4	30.8	38506	9	US-09-793-708-19	Sequence 19, Appl
10	74.6	28.9	13842	9	US-09-860-846-30	Sequence 30, Appl
11	74.6	28.9	13842	9	US-09-988-384B-30	Sequence 30, Appl
12	74.6	28.9	13842	9	US-09-836-821-30	Sequence 30, Appl
13	74.6	28.9	13842	10	US-09-861-289-30	Sequence 30, Appl
14	73.6	28.5	4851	9	US-09-712-363-116	Sequence 116, App
15	73.2	28.4	15872	9	US-09-860-846-1	Sequence 1, Appli
16	73.2	28.4	15872	9	US-09-988-384B-1	Sequence 1, Appli
17	73.2	28.4	15872	9	US-09-836-821-1	Sequence 1, Appli
18	73.2	28.4	15872	10	US-09-861-289-1	Sequence 1, Appli
19	69	26.7	50937	9	US-09-808-880-1	Sequence 1, Appli
20	68.4	26.5	4689	9	US-09-860-846-34	Sequence 34, Appl
21	68.4	26.5	4689	9	US-09-988-384B-34	Sequence 34, Appl
22	68.4	26.5	4689	9	US-09-836-821-34	Sequence 34, Appl
23	68.4	26.5	4689	10	US-09-861-289-34	Sequence 34, Appl
24	58.6	22.7	4041	9	US-09-860-846-36	Sequence 36, Appl
25	58.6	22.7	4041	9	US-09-988-384B-36	Sequence 36, Appl
26	58.6	22.7	4041	9	US-09-836-821-36	Sequence 36, Appl
27	58.6	22.7	4041	10	US-09-861-289-36	Sequence 36, Appl
28	55	21.3	68750	9	US-10-014-717-1	Sequence 1, Appli
29	48	18.6	2305	10	US-09-073-009-2	Sequence 2, Appli
30	48	18.6	2305	10	US-09-023-588-2	Sequence 2, Appli
31	48	18.6	2305	10	US-09-793-306-2	Sequence 2, Appli
32	47	18.2	88421	9	US-09-976-059-1	Sequence 1, Appli
33	42.8	16.6	1592	9	US-09-954-531-1363	Sequence 1363, Ap
34	42.8	16.6	1928	9	US-10-198-846-9834	Sequence 9834, Ap
35	41.6	16.1	1074	10	US-09-815-242-7911	Sequence 7911, Ap
36	41.4	16.0	4209	9	US-09-712-363-20	Sequence 20, Appl
c 37	40.4	15.7	38734	9	US-09-373-658-30	Sequence 30, Appl
38	40.2	15.6	597	9	US-09-894-844-11	Sequence 11, Appl
39	39.8	15.4	893	9	US-10-037-270-935	Sequence 935, App
40	39.2	15.2	1248	9	US-09-860-846-7	Sequence 7, Appli
41	39.2	15.2	1248	9	US-09-988-384B-7	Sequence 7, Appli
42	39.2	15.2	1248	9	US-09-836-821-7	Sequence 7, Appli
43	39.2	15.2	1248	10	US-09-861-289-7	Sequence 7, Appli
c 44	39.2	15.2	12441	9	US-09-988-384B-3	Sequence 3, Appli
c 45	39.2	15.2	13613	9	US-09-860-846-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-860-846-32
; Sequence 32, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-32

Query Match 30.8%; Score 79.4; DB 9; Length 11220;
Best Local Similarity 60.2%; Pred. No. 1.6e-13;
Matches 150; Conservative 0; Mismatches 96; Indels 3; Gaps 1;

```
Qy          9 GCTGTCCCCCACGCAGCAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGC 68
              ||| || | | |  | | |   | | ||  | |||||  | | | | |  | |
Db      10677 GCTCGCCGCGATGACACCGGACGACCGGGTCGCGCACCTGCGGGACCTCGTCCGTACGCA 10736

Qy         69 GGTGCTGAACGACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTT 128
               |||  || || |   ||| |||   ||||| ||       | | ||    | |  ||
Db      10737 CGTGGCGACCGTCCTGGGACACG---GCACCCCGAGCCGGGTGGACCTGGAGCGGGCCTT 10793

Qy        129 CGCCCACCTCGGCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGC 188
              |  | ||  ||| |||||||||  |||  |  |||||||| ||||||| |||| |   ||
Db      10794 CCGCGACACCGGTTTCGACTCGCTCACCGCCGTCGAACTCCGCAACCGTCTCAACGCCGC 10853

Qy        189 CACGGGCCTGCGGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGC 248
              || || |||||| |||| |  |||||  |||||||||||| |||| ||||  | | ||||
Db      10854 GACCGGGCTGCGGCTGCCGGCCACGCTGGTCTTCGACCACCCCACCCCGGGGGAGCTCGC 10913
```

```
Db      12601 GCCGTCTTCGGTGGCGCGGGGCAGGGCGCCTACGCCGCCGCCAACGCCACCCTCGACGCC 12660

Qy        475 CTGGCGTACCGGCGGCGGGCGGCGGGTCTGCCGGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534

Db      12661 CTCGCCTGGCGCCGCCGGGCAGCCGGACTCCCCGCCCTCTCCCTCGGCTGGGGCCTCTGG 12720


RESULT 15
US-09-860-846-32
; Sequence 32, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
;   LENGTH: 11220
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-32

  Query Match            36.8%;  Score 196.4;  DB 9;  Length 11220;
  Best Local Similarity  61.7%;  Pred. No. 2.1e-38;
  Matches  333;  Conservative    0;  Mismatches  201;  Indels    6;  Gaps    1;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60

Db       9829 GGTACGGTCCTGCTGACCGGCGGCACCGGTGCGCTGGGGGGCATCGTGGCCCGGCACGTG 9888

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120

Db       9889 GTGGGCGAGTGGGGCGTACGACGCCTGCTGCTCGTGAGCCGGCGGGGCACGGACGCCCCG 9948

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180

Db       9949 GGCGCCGGCGAGCTCGTGCACGAGCTGGAGGCCCTGGGAGCCGACGTCTCGGTGGCCGCG 10008

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240

Db      10009 TGCGACGTCGCCGACCGCGAAGCCCTCACCGCCGTACTCGACTCGATCCCCGCCGAACAC 10068

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300

Db      10069 CCGCTCACCGCGGTCGTCCACACGGCAGGCGTCCTCTCCGACGGCACCCTCCCCTCGATG 10128

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360

Db      10129 ACAGCGGAGGATGTGGAACACGTACTGCGTCCCAAGGTCGACGCCGCGTTCCTCCTCGAC 10188

Qy        361 GAGCT------GACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCG 414

Db      10189 GAACTCACCTCGACGCCCGGCTACGACCTGGCAGCGTTCGTCATGTTCTCCTCCGCCGCC 10248

Qy        415 GGGATCCTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCG 474

Db      10249 GCCGTCTTCGGTGGCGCGGGGCAGGGCGCCTACGCCGCCGCCAACGCCACCCTCGACGCC 10308

Qy        475 CTGGCGTACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534

Db      10309 CTCGCCTGGCGCCGCCGGACAGCCGGACTCCCCGCCCTCTCCCTCGGCTGGGGCCTCTGG 10368


Search completed: June 18, 2003, 05:08:46
```

Job time : 66.1981 secs

```
Qy        361 GAGCT------GACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCG 414
              ||||       |||||  |      || ||| | ||||||||  |||||||||| |||||
Db      14399 GAACTCACCTCGACGCCCGCATACGACCTGGCAGCGTTCGTCATGTTCTCCTCCGCCGCC 14458

Qy        415 GGGATCCTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCG 474
              |   || | ||    || |||||||||  |||||||| ||||| ||| | || |||||
Db      14459 GCCGTCTTCGGTGGCGCGGGGCAGGGCGCCTACGCCGCCGCCAACGCCACCCTCGACGCC 14518

Qy        475 CTGGCGTACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              || || |  || || ||||| || || || || |   | || || |  ||||| || |||
Db      14519 CTCGCCTGGCGCCGCCGGGCAGCCGGACTCCCCGCCCTCTCCCTCGGCTGGGGCCTCTGG 14578


RESULT 13
US-09-988-384B-5
; Sequence 5, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 5
;   LENGTH: 37948
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-5

  Query Match             37.4%;  Score 199.6;  DB 9;  Length 37948;
  Best Local Similarity   62.0%;  Pred. No. 3.2e-39;
  Matches  335;  Conservative    0;  Mismatches  199;  Indels    6;  Gaps    1;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||  || |||| || || || || |  ||||| |    ||| || |||||  ||
Db      15209 GGCACGGTCCTGCTGACCGGCGGCACCGGCGGCCTGGGCGGACTGGTCGCCCGGCACGTG 15268

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  || | ||| || || ||     |||||| |||||||||||||| ||  |||| || |||
Db      15269 GTGGGCGAGTGGGGCGTACGACGCCTGCTGCTGGTGAGCCGGCGGGGCACGGACGCCCCG 15328

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || || || |  ||   |  |||||||  ||  |||| || || ||   | | |  |||
Db      15329 GGCGCCGACGAGCTCGTGCACGAGCTGGAGGCCCTGGGAGCCGACGTCTCGGTGGCCGCG 15388

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              || || || |  || ||    |   |   |  || || |   | || ||||
Db      15389 TGCGACGTCGCCGACCGCGAAGCCCTCACCGCCGTACTCGACGCCATCCCCGCCGAACAC 15448

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||| || |  |||||  ||  |||| || || ||    ||| |  ||  || | ||  |
Db      15449 CCGCTCACCGCGGTCGTCCACACGGCAGGCGTCCTCTCCGACGGCACCCTCCCGTCCATG 15508

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||| | |||    |||     ||  |   | |||||| || |||||  |  | || ||
Db      15509 ACGACGGAGGACGTGGAACACGTACTGCGGCCCAAGGTCGACGCCGCGTTCCTCCTCGAC 15568

Qy        361 GAGCT------GACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCG 414
              || ||      |||||  |      || ||| | ||||||||  |||||||||| |||||
Db      15569 GAACTCACCTCGACGCCCGCATACGACCTGGCAGCGTTCGTCATGTTCTCCTCCGCCGCC 15628

Qy        415 GGGATCCTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCG 474
              |   || | ||    || |||||||||  |||||||| ||||| ||| | || |||||
Db      15629 GCCGTCTTCGGTGGCGCGGGGCAGGGCGCCTACGCCGCCGCCAACGCCACCCTCGACGCC 15688

Qy        475 CTGGCGTACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              || || |  || || ||||| || || || || |   | || || |  ||||| || |||
Db      15689 CTCGCCTGGCGCCGCCGGGCAGCCGGACTCCCCGCCCTCTCCCTCGGCTGGGGCCTCTGG 15748


RESULT 14
US-09-793-708-19
; Sequence 19, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT:  BETLACH, Melanie C.
;  APPLICANT:  BETLACH, Mary C.
;  APPLICANT:  McDANIEL, Robert
;  APPLICANT:  TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002121
;  CURRENT APPLICATION NUMBER: US/09/793,708
;  CURRENT FILING DATE:  2001-02-22
;  PRIOR APPLICATION NUMBER: US 09/657,440
;  PRIOR FILING DATE: 2000-09-07
;  PRIOR APPLICATION NUMBER: US 09/320,878
;  PRIOR FILING DATE: 1999-05-27
;  PRIOR APPLICATION NUMBER: US 09/141,908
;  PRIOR FILING DATE: 1998-08-28
;  PRIOR APPLICATION NUMBER: US 09/073,538
;  PRIOR FILING DATE: 1998-05-06
;  PRIOR APPLICATION NUMBER: US 08/846,247
;  PRIOR FILING DATE: 1997-04-30
;  PRIOR APPLICATION NUMBER: US 60/134,990
;  PRIOR FILING DATE: 1999-05-20
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-793-708-19

  Query Match             37.4%;  Score 199.6;  DB 9;  Length 38506;
  Best Local Similarity   62.0%;  Pred. No. 3.2e-39;
  Matches  335;  Conservative    0;  Mismatches  199;  Indels    6;  Gaps    1;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||  || |||| || || || || |  ||||| |    ||| || |||||  ||
Db      12181 GGCACGGTCCTGCTGACCGGCGGCACCGGCGGCCTGGGCGGACTGGTCGCCCGGCACGTG 12240

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  || | ||| || || ||     |||||| |||||||||||||| ||  |||| || |||
Db      12241 GTGGGCGAGTGGGGCGTACGACGCCTGCTGCTGGTGAGCCGGCGGGGCACGGACGCCCCG 12300

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || || || |  ||   |  |||||||  ||  |||| || || ||   | | |  |||
Db      12301 GGCGCCGACGAGCTCGTGCACGAGCTGGAGGCCCTGGGAGCCGACGTCTCGGTGGCCGCG 12360

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              || || || |  || ||    |   |   |  || || |   | || ||||
Db      12361 TGCGACGTCGCCGACCGCGAAGCCCTCACCGCCGTACTCGACGCCATCCCCGCCGAACAC 12420

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||| || |  |||||  ||  |||| || || ||    ||| |  ||  || | ||  |
Db      12421 CCGCTCACCGCGGTCGTCCACACGGCAGGCGTCCTCTCCGACGGCACCCTCCCGTCCATG 12480

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||| | |||    |||     ||  |   | |||||| || |||||  |  | || ||
Db      12481 ACGACGGAGGACGTGGAACACGTACTGCGGCCCAAGGTCGACGCCGCGTTCCTCCTCGAC 12540

Qy        361 GAGCT------GACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCG 414
              || ||      |||||  |      || ||| | ||||||||  |||||||||| |||||
Db      12541 GAACTCACCTCGACGCCCGCATACGACCTGGCAGCGTTCGTCATGTTCTCCTCCGCCGCC 12600

Qy        415 GGGATCCTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCG 474
```

```
Db      14099 GTGGGCGAGTGGGGCGTACGACGCCTGCTGCTGGTGAGCCGGCGGGGCACGGACGCCCCG 14158

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180

Db      14159 GGCGCCGACGAGCTCGTGCACGAGCTGGAGGCCCTGGGAGCCGACGTCTCGGTGGCCGCG 14218

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240

Db      14219 TGCGACGTCGCCGACCGCGAAGCCCTCACCGCCGTACTCGACGCCATCCCCGCCGAACAC 14278

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300

Db      14279 CCGCTCACCGCGGTCGTCCACACGGCAGGCGTCCTCTCCGACGGCACCCTCCCGTCCATG 14338

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360

Db      14339 ACGACGGAGGACGTGGAACACGTACTGCGGCCCAAGGTCGACGCCGCGTTCCTCCTCGAC 14398

Qy        361 GAGCT------GACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCG 414

Db      14399 GAACTCACCTCGACGCCCGCATACGACCTGGCAGCGTTCGTCATGTTCTCCTCCGCCGCC 14458

Qy        415 GGGATCCTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCG 474

Db      14459 GCCGTCTTCGGTGGCGCGGGGCAGGGCGCCTACGCCGCCGCCAACGCCACCCTCGACGCC 14518

Qy        475 CTGGCGTACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534

Db      14519 CTCGCCTGGCGCCGCCGGGCAGCCGGACTCCCCGCCCTCTCCCTCGGCTGGGGCCTCTGG 14578
```

RESULT 11
US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-5

Query Match 37.4%; Score 199.6; DB 9; Length 36778;
Best Local Similarity 62.0%; Pred. No. 3.2e-39;
Matches 335; Conservative 0; Mismatches 199; Indels 6; Gaps 1;

```
Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60

Db      14039 GGCACGGTCCTGCTGACCGGCGGCACCGGCGGCCTGGGCGGACTGGTCGCCCGGCACGTG 14098

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120

Db      14099 GTGGGCGAGTGGGGCGTACGACGCCTGCTGCTGGTGAGCCGGCGGGGCACGGACGCCCCG 14158

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180

Db      14159 GGCGCCGACGAGCTCGTGCACGAGCTGGAGGCCCTGGGAGCCGACGTCTCGGTGGCCGCG 14218

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240

Db      14219 TGCGACGTCGCCGACCGCGAAGCCCTCACCGCCGTACTCGACGCCATCCCCGCCGAACAC 14278

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300

Db      14279 CCGCTCACCGCGGTCGTCCACACGGCAGGCGTCCTCTCCGACGGCACCCTCCCGTCCATG 14338

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360

Db      14339 ACGACGGAGGACGTGGAACACGTACTGCGGCCCAAGGTCGACGCCGCGTTCCTCCTCGAC 14398

Qy        361 GAGCT------GACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCG 414

Db      14399 GAACTCACCTCGACGCCCGCATACGACCTGGCAGCGTTCGTCATGTTCTCCTCCGCCGCC 14458

Qy        415 GGGATCCTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCG 474

Db      14459 GCCGTCTTCGGTGGCGCGGGGCAGGGCGCCTACGCCGCCGCCAACGCCACCCTCGACGCC 14518

Qy        475 CTGGCGTACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534

Db      14519 CTCGCCTGGCGCCGCCGGGCAGCCGGACTCCCCGCCCTCTCCCTCGGCTGGGGCCTCTGG 14578
```

RESULT 12
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5

Query Match 37.4%; Score 199.6; DB 10; Length 36778;
Best Local Similarity 62.0%; Pred. No. 3.2e-39;
Matches 335; Conservative 0; Mismatches 199; Indels 6; Gaps 1;

```
Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60

Db      14039 GGCACGGTCCTGCTGACCGGCGGCACCGGCGGCCTGGGCGGACTGGTCGCCCGGCACGTG 14098

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120

Db      14099 GTGGGCGAGTGGGGCGTACGACGCCTGCTGCTGGTGAGCCGGCGGGGCACGGACGCCCCG 14158

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180

Db      14159 GGCGCCGACGAGCTCGTGCACGAGCTGGAGGCCCTGGGAGCCGACGTCTCGGTGGCCGCG 14218

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240

Db      14219 TGCGACGTCGCCGACCGCGAAGCCCTCACCGCCGTACTCGACGCCATCCCCGCCGAACAC 14278

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300

Db      14279 CCGCTCACCGCGGTCGTCCACACGGCAGGCGTCCTCTCCGACGGCACCCTCCCGTCCATG 14338

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360

Db      14339 ACGACGGAGGACGTGGAACACGTACTGCGGCCCAAGGTCGACGCCGCGTTCCTCCTCGAC 14398
```

```
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-30

  Query Match             37.4%;  Score 199.6;  DB 9;  Length 13842;
  Best Local Similarity   62.0%;  Pred. No. 3.6e-39;
  Matches  335;  Conservative    0;  Mismatches  199;  Indels    6;  Gaps    1;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||  || ||||  || ||  ||  ||  |  |||||  |     |||  ||  |||||  ||
Db      12298 GGCACGGTCCTGCTGACCGGCGGCACCGGCGGCCTGGGCGGACTGGTCGCCCGGCACGTG 12357

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  ||  |  |||  ||  ||  ||     ||||||  ||||||||||||||  ||   ||||  ||  |||
Db      12358 GTGGGCGAGTGGGGCGTACGACGCCTGCTGCTGGTGAGCCGGCGGGGCACGGACGCCCCG 12417

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              ||  ||  ||  |   ||    |   |||||||  ||  ||||  ||  ||  ||    |  |  |||
Db      12418 GGCGCCGACGAGCTCGTGCACGAGCTGGAGGCCCTGGGAGCCGACGTCTCGGTGGCCGCG 12477

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||  ||  ||  |  ||  ||     |   |     |  ||  ||  |    |  ||  ||||
Db      12478 TGCGACGTCGCCGACCGCGAAGCCCTCACCGCCGTACTCGACGCCATCCCCGCCGAACAC 12537

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |||||  ||  |  |||||  ||  ||||  ||  ||  ||     |||  |   ||  ||  |  ||  |
Db      12538 CCGCTCACCGCGGTCGTCCACACGGCAGGCGTCCTCTCCGACGGCACCCTCCCGTCCATG 12597

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |||  |  |||     |||      ||  |    |  |||||||  ||  |||||    |   |  ||  ||
Db      12598 ACGACGGAGGACGTGGAACACGTACTGCGGCCCAAGGTCGACGCCGCGTTCCTCCTCGAC 12657

Qy        361 GAGCT------GACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCG 414
              ||  ||       |||||  |       ||  |||  |  ||||||||   ||||||||||  |||||
Db      12658 GAACTCACCTCGACGCCCGCATACGACCTGGCAGCGTTCGTCATGTTCTCCTCCGCCGCC 12717

Qy        415 GGGATCCTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCG 474
              |    ||  |  ||     ||  |||||||||   |||||||||  |||||  |||  |  ||  |||||
Db      12718 GCCGTCTTCGGTGGCGCGGGGCAGGGCGCCTACGCCGCCGCCAACGCCACCCTCGACGCC 12777

Qy        475 CTGGCGTACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||  ||  |   ||  ||  |||||  ||  ||  ||  ||  |     |  ||  ||  |  |||||  ||  |||
Db      12778 CTCGCCTGGCGCCGCCGGGCAGCCGGACTCCCCGCCCTCTCCCTCGGCTGGGGCCTCTGG 12837


RESULT 9
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-30

  Query Match             37.4%;  Score 199.6;  DB 10;  Length 13842;
  Best Local Similarity   62.0%;  Pred. No. 3.6e-39;
  Matches  335;  Conservative    0;  Mismatches  199;  Indels    6;  Gaps    1;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||  || ||||  || ||  ||  ||  |  |||||  |     |||  ||  |||||  ||
Db      12298 GGCACGGTCCTGCTGACCGGCGGCACCGGCGGCCTGGGCGGACTGGTCGCCCGGCACGTG 12357

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  ||  |  |||  ||  ||  ||     ||||||  ||||||||||||||  ||   ||||  ||  |||
Db      12358 GTGGGCGAGTGGGGCGTACGACGCCTGCTGCTGGTGAGCCGGCGGGGCACGGACGCCCCG 12417

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              ||  ||  ||  |   ||    |   |||||||  ||  ||||  ||  ||  ||    |  |  |||
Db      12418 GGCGCCGACGAGCTCGTGCACGAGCTGGAGGCCCTGGGAGCCGACGTCTCGGTGGCCGCG 12477

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||  ||  ||  |  ||  ||     |   |     |  ||  ||  |    |  ||  ||||
Db      12478 TGCGACGTCGCCGACCGCGAAGCCCTCACCGCCGTACTCGACGCCATCCCCGCCGAACAC 12537

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |||||  ||  |  |||||  ||  ||||  ||  ||  ||     |||  |   ||  ||  |  ||  |
Db      12538 CCGCTCACCGCGGTCGTCCACACGGCAGGCGTCCTCTCCGACGGCACCCTCCCGTCCATG 12597

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |||  |  |||     |||      ||  |    |  |||||||  ||  |||||    |   |  ||  ||
Db      12598 ACGACGGAGGACGTGGAACACGTACTGCGGCCCAAGGTCGACGCCGCGTTCCTCCTCGAC 12657

Qy        361 GAGCT------GACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCG 414
              ||  ||       |||||  |       ||  |||  |  ||||||||   ||||||||||  |||||
Db      12658 GAACTCACCTCGACGCCCGCATACGACCTGGCAGCGTTCGTCATGTTCTCCTCCGCCGCC 12717

Qy        415 GGGATCCTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCG 474
              |    ||  |  ||     ||  |||||||||   |||||||||  |||||  |||  |  ||  |||||
Db      12718 GCCGTCTTCGGTGGCGCGGGGCAGGGCGCCTACGCCGCCGCCAACGCCACCCTCGACGCC 12777

Qy        475 CTGGCGTACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||  ||  |   ||  ||  |||||  ||  ||  ||  ||  |     |  ||  ||  |  |||||  ||  |||
Db      12778 CTCGCCTGGCGCCGCCGGGCAGCCGGACTCCCCGCCCTCTCCCTCGGCTGGGGCCTCTGG 12837


RESULT 10
US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-5

  Query Match             37.4%;  Score 199.6;  DB 9;  Length 36778;
  Best Local Similarity   62.0%;  Pred. No. 3.2e-39;
  Matches  335;  Conservative    0;  Mismatches  199;  Indels    6;  Gaps    1;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||  || ||||  || ||  ||  ||  |  |||||  |     |||  ||  |||||  ||
Db      14039 GGCACGGTCCTGCTGACCGGCGGCACCGGCGGCCTGGGCGGACTGGTCGCCCGGCACGTG 14098

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
```

```
Db       4162 GCCGCTCACCGGCAGGCCGCAGGGTTGGCCGGGATATCACTGGCGTGGGGTTTGTGG 4218


RESULT 6
US-09-860-846-30
; Sequence 30, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-30

  Query Match             37.4%;  Score 199.6;  DB 9;  Length 13842;
  Best Local Similarity   62.0%;  Pred. No. 3.6e-39;
  Matches  335;  Conservative    0;  Mismatches  199;  Indels    6;  Gaps    1;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||  || |||| || || || || | ||||| |     ||| || |||||  ||
Db      12298 GGCACGGTCCTGCTGACCGGCGGCACCGGCGGCCTGGGCGGACTGGTCGCCCGGCACGTG 12357

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  || | ||| || || ||    |||||| |||||||||||||| ||  |||| || |||
Db      12358 GTGGGCGAGTGGGGCGTACGACGCCTGCTGCTGGTGAGCCGGCGGGGCACGGACGCCCCG 12417

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || || || |  ||  |  |||||||  ||  |||| || || ||   | | |  |||
Db      12418 GGCGCCGACGAGCTCGTGCACGAGCTGGAGGCCCTGGGAGCCGACGTCTCGGTGGCCGCG 12477

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              || || || |  || ||    |  |  |    | || || |   | || ||||
Db      12478 TGCGACGTCGCCGACCGCGAAGCCCTCACCGCCGTACTCGACGCCATCCCCGCCGAACAC 12537

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||| || |  ||||| ||  |||| || || ||   ||| |  ||  || | ||  |
Db      12538 CCGCTCACCGCGGTCGTCCACACGGCAGGCGTCCTCTCCGACGGCACCCTCCCGTCCATG 12597

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||| | |||    |||     ||  |   | |||||||| || |||||  |  | || ||
Db      12598 ACGACGGAGGACGTGGAACACGTACTGCGGCCCAAGGTCGACGCCGCGTTCCTCCTCGAC 12657

Qy        361 GAGCT------GACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCG 414
              || ||      |||||  |      || ||| | |||||||| ||||||||| |||||
Db      12658 GAACTCACCTCGACGCCCGCATACGACCTGGCAGCGTTCGTCATGTTCTCCTCCGCCGCC 12717

Qy        415 GGGATCCTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCG 474
              |   || | ||    || |||||||| |||||||| |||| ||| | || |||||
Db      12718 GCCGTCTTCGGTGGCGCGGGGCAGGGCGCCTACGCCGCCGCCAACGCCACCCTCGACGCC 12777

Qy        475 CTGGCGTACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              || || |  || || |||||  || || || || |   | || || |  |||| || |||
Db      12778 CTCGCCTGGCGCCGCCGGGCAGCCGGACTCCCCGCCCTCTCCCTCGGCTGGGGCCTCTGG 12837


RESULT 7
US-09-988-384B-30
; Sequence 30, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-30

  Query Match             37.4%;  Score 199.6;  DB 9;  Length 13842;
  Best Local Similarity   62.0%;  Pred. No. 3.6e-39;
  Matches  335;  Conservative    0;  Mismatches  199;  Indels    6;  Gaps    1;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||  || |||| || || || || | ||||| |     ||| || |||||  ||
Db      12298 GGCACGGTCCTGCTGACCGGCGGCACCGGCGGCCTGGGCGGACTGGTCGCCCGGCACGTG 12357

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  || | ||| || || ||    |||||| |||||||||||||| ||  |||| || |||
Db      12358 GTGGGCGAGTGGGGCGTACGACGCCTGCTGCTGGTGAGCCGGCGGGGCACGGACGCCCCG 12417

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || || || |  ||  |  |||||||  ||  |||| || || ||   | | |  |||
Db      12418 GGCGCCGACGAGCTCGTGCACGAGCTGGAGGCCCTGGGAGCCGACGTCTCGGTGGCCGCG 12477

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              || || || |  || ||    |  |  |    | || || |   | || ||||
Db      12478 TGCGACGTCGCCGACCGCGAAGCCCTCACCGCCGTACTCGACGCCATCCCCGCCGAACAC 12537

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||| || |  ||||| ||  |||| || || ||   ||| |  ||  || | ||  |
Db      12538 CCGCTCACCGCGGTCGTCCACACGGCAGGCGTCCTCTCCGACGGCACCCTCCCGTCCATG 12597

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||| | |||    |||     ||  |   | |||||||| || |||||  |  | || ||
Db      12598 ACGACGGAGGACGTGGAACACGTACTGCGGCCCAAGGTCGACGCCGCGTTCCTCCTCGAC 12657

Qy        361 GAGCT------GACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCG 414
              || ||      |||||  |      || ||| | |||||||| ||||||||| |||||
Db      12658 GAACTCACCTCGACGCCCGCATACGACCTGGCAGCGTTCGTCATGTTCTCCTCCGCCGCC 12717

Qy        415 GGGATCCTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCG 474
              |   || | ||    || |||||||| |||||||| |||| ||| | || |||||
Db      12718 GCCGTCTTCGGTGGCGCGGGGCAGGGCGCCTACGCCGCCGCCAACGCCACCCTCGACGCC 12777

Qy        475 CTGGCGTACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              || || |  || || |||||  || || || || |   | || || |  |||| || |||
Db      12778 CTCGCCTGGCGCCGCCGGGCAGCCGGACTCCCCGCCCTCTCCCTCGGCTGGGGCCTCTGG 12837


RESULT 8
US-09-836-821-30
; Sequence 30, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
```

```
Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||  ||    ||  |||| ||| | |    ||||||||| ||||| || |||
Db       7817 GCCCACCGCGCCGCCACGGGCCTGGCCGCCACGTCGCTGGCCTGGGGCCTCTGG 7870


RESULT 4
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-1

  Query Match             39.8%;  Score 212.4;  DB 10;  Length 15872;
  Best Local Similarity   62.4%;  Pred. No. 3.1e-42;
  Matches  333;  Conservative    0;  Mismatches  201;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  | ||  | || || || || || || || |  || || |||    | || || || ||
Db       7337 GGCACCGTCCTCGTCACCGGGGGCACCGGAGGCCTCGGAGCGCTCCTCGCCCGCCACCTC 7396

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  |     ||||||||    |||||| ||||||||| || ||  | || || ||
Db       7397 GTGGAGCGTCACGGGGTGCGCCGGCTGCTGCTGGTGAGCCGCCGCGGGGCCGACGCCCCC 7456

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || ||||  |  ||| | | ||| || || |   | || |||||||||  | | |  ||
Db       7457 GGCGCGGCCGACCTGGGCGAGGACCTCGCGGGCCTCGGCGCGGAGGTGGCGTTCGCCGCC 7516

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              || |  |  || ||   |   ||| ||| | | |      || || |||| |  |
Db       7517 GCCGACGCCGCCGACCGCGAGAGCCTGGCGCGGGCGATCGCCACCGTGCCCGCCGAGCAT 7576

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||||  |||||||  |||| || ||  | ||||| ||||||  | |   | |||
Db       7577 CCGCTGACGGCCGTCGTGCACACGGCGGGAGTCGTCGACGACGCGACGGTGGAGGCGCTC 7636

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              || || || ||||||| | ||||  |     | ||||| || |||||        ||| |
Db       7637 ACACCGGAACGGCTGGACGCGGTACTGCGCCCGAAGGTCGACGCCGCGTGGAACCTGCAC 7696

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||||| ||   ||   ||  |||    || ||||| || |||||||| | | | || |||
Db       7697 GAGCTCACCAAGGACCTGCGGCTCGACGCCTTCGTCCTCTTCTCCTCCGTCTCCGGCATC 7756

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | ||  | ||||| |||| |||||||| ||||||||  | |  || ||||| || ||
Db       7757 GTCGGCACCGCCGGCCAGGCCAACTACGCGGCGGCCAACACGGGCCTCGACGCCCTCGCC 7816

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||  ||    ||  |||| ||| | |    ||||||||| ||||| || |||
Db       7817 GCCCACCGCGCCGCCACGGGCCTGGCCGCCACGTCGCTGGCCTGGGGCCTCTGG 7870


RESULT 5
US-09-712-363-116
; Sequence 116, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
;  APPLICANT: Eisenberg, David
;  APPLICANT:  Rotstein, Sergio H.
;  APPLICANT:  Marcotte, Edward M.
;  TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
;  TITLE OF INVENTION:  INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
;  FILE REFERENCE: 07419-032001
;  CURRENT APPLICATION NUMBER: US/09/712,363
;  CURRENT FILING DATE:  2000-11-13
;  PRIOR APPLICATION NUMBER: PCT/US00/02246
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: 60/179,531
;  PRIOR FILING DATE: 2000-02-01
;  PRIOR APPLICATION NUMBER: 60/117,844
;  PRIOR FILING DATE: 1999-01-29
;  PRIOR APPLICATION NUMBER: 60/118,206,
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: 60/126,593
;  PRIOR FILING DATE: 1999-03-26
;  PRIOR APPLICATION NUMBER: 60/134,093
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: 60/134,092
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: 60/165,124
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/165,086
;  PRIOR FILING DATE: 1999-11-12
;  NUMBER OF SEQ ID NOS: 292
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 116
;   LENGTH: 4851
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
US-09-712-363-116

  Query Match             38.8%;  Score 207.4;  DB 9;  Length 4851;
  Best Local Similarity   62.9%;  Pred. No. 5.5e-41;
  Matches  338;  Conservative    0;  Mismatches  196;  Indels    3;  Gaps    1;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||| || | || |||| || || | | |||| || | |||| || || |||
Db       3682 GGCACGGTGGTGATCACCGGTGCCACCGGGGCGGTTGGTGGGGTGTTGGCCCGCCACCTG 3741

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              | ||| |  | |||||||||  ||||| ||||||  || |||| ||         ||   |
Db       3742 GTTGGCGCCTATGGGGTGCGTCATCTGGTGTTGGCCAGTCGGCGGGGCGATCGCGCGGAG 3801

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || ||||  |   ||  || ||  || | | |   || ||  |||||  | | || ||
Db       3802 GGAGCGGCCGAATTGGCCGCCGACTTGACGGAGGCCGGCGCCAAGGTGCAGGTGGTGGCC 3861

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||| |||| || ||  | ||| |   || || | |       |  |    | ||
Db       3862 TGTGACGTGGCCGATCGCGCTGCGGTAGCGGGGTTGTTTGCCCAGCTGTCGCGGGAGTAC 3921

Qy        241 CCGC---TGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCT 297
              ||||   ||   || ||  | ||||| || || ||||| || || ||   |||| ||||
Db       3922 CCGCCGGTGCGCGGGGTGATTCATGCCGCCGGCGTGCTCGATGACGCAGTGATCACCTCG 3981

Qy        298 CTCACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTG 357
              | || || || ||  | |  |||||||   |||||||||||| || ||      |||
Db       3982 TTGACACCGGACCGCATCGATACGGTGTTGCGGGCCAAGGTGGACGCGGCGTGGAACCTG 4041

Qy        358 GATGAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGG 417
              |  ||   ||  | |  ||||  |||||  ||| | ||||| ||| |||  |||||
Db       4042 CACCAGGCCACCAGTGACCTGGATTTGTCGATGTTTGCGCTGTGCTCATCGATCGCGGCC 4101

Qy        418 ATCCTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTG 477
              |   | || ||| | ||||||||||||||| | ||||| || ||   |||||||| | ||
Db       4102 ACGGTCGGCTCGCCGGGGCAGGGCAACTACTCGGCGGCAAACGCGTTTCTGGACGGGTTG 4161

Qy        478 GCGTACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||    | |||| ||| || ||  || | ||| | || ||||||||||  |||||
```

```
Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||||  ||||||||  |||| || ||  | |||||  ||||||  | |    | |||
Db       7577 CCGCTGACGGCCGTCGTGCACACGGCGGGAGTCGTCGACGACGCGACGGTGGAGGCGCTC 7636

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              || || ||  |||||||  | ||||   |      | |||||  || |||||       ||| |
Db       7637 ACACCGGAACGGCTGGACGCGGTACTGCGCCCGAAGGTCGACGCCGCGTGGAACCTGCAC 7696

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||||  ||   ||   ||  |||     || |||||  || ||||||| | |  || |||
Db       7697 GAGCTCACCAAGGACCTGCGGCTCGACGCCTTCGTCCTCTTCTCCTCCGTCTCCGGCATC 7756

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | ||  | |||||  ||||  |||||||| ||||||||  | |   || |||||  || ||
Db       7757 GTCGGCACCGCCGGCCAGGCCAACTACGCGGCGGCCAACACGGGCCTCGACGCCCTCGCC 7816

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                ||  ||    ||  |||| ||| | |    |||||||| |||||  || |||
Db       7817 GCCCACCGCGCCGCCACGGGCCTGGCCGCCACGTCGCTGGCCTGGGGCCTCTGG 7870
```

RESULT 2
US-09-988-384B-1
; Sequence 1, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-1

Query Match 39.8%; Score 212.4; DB 9; Length 15872;
Best Local Similarity 62.4%; Pred. No. 3.1e-42;
Matches 333; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

```
Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  | ||  | || || || || || || |  || || |||    | || || || ||
Db       7337 GGCACCGTCCTCGTCACCGGGGGCACCGGAGGCCTCGGAGCGCTCCTCGCCCGCCACCTC 7396

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              | |          |||||||     |||||| ||||||||| || ||  | || || ||
Db       7397 GTGGAGCGTCACGGGGTGCGCCGGCTGCTGCTGGTGAGCCGCCGCGGGGCCGACGCCCCC 7456

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || |||| |  ||| | | ||| || || |   | || |||||||||  | | |  ||
Db       7457 GGCGCGGCCGACCTGGGCGAGGACCTCGCGGGCCTCGGCGCGGAGGTGGCGTTCGCCGCC 7516

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              || |  |  || ||   |   |||  |||  | | |     || || |||| | |
Db       7517 GCCGACGCCGCCGACCGCGAGAGCCTGGCGCGGGCGATCGCCACCGTGCCCGCCGAGCAT 7576

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |||||||||  |||||||  |||| || ||  | |||||  ||||||  | |    | |||
Db       7577 CCGCTGACGGCCGTCGTGCACACGGCGGGAGTCGTCGACGACGCGACGGTGGAGGCGCTC 7636

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              || || ||  |||||||  | ||||   |      | |||||  || |||||       ||| |
Db       7637 ACACCGGAACGGCTGGACGCGGTACTGCGCCCGAAGGTCGACGCCGCGTGGAACCTGCAC 7696
```

```
Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||||  ||   ||   ||  |||     || |||||  || ||||||| | |  || |||
Db       7697 GAGCTCACCAAGGACCTGCGGCTCGACGCCTTCGTCCTCTTCTCCTCCGTCTCCGGCATC 7756

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | ||  | |||||  ||||  |||||||| ||||||||  | |   || |||||  || ||
Db       7757 GTCGGCACCGCCGGCCAGGCCAACTACGCGGCGGCCAACACGGGCCTCGACGCCCTCGCC 7816

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                ||  ||    ||  |||| ||| | |    |||||||| |||||  || |||
Db       7817 GCCCACCGCGCCGCCACGGGCCTGGCCGCCACGTCGCTGGCCTGGGGCCTCTGG 7870
```

RESULT 3
US-09-836-821-1
; Sequence 1, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-1

Query Match 39.8%; Score 212.4; DB 9; Length 15872;
Best Local Similarity 62.4%; Pred. No. 3.1e-42;
Matches 333; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

```
Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  | ||  | || || || || || || |  || || |||    | || || || ||
Db       7337 GGCACCGTCCTCGTCACCGGGGGCACCGGAGGCCTCGGAGCGCTCCTCGCCCGCCACCTC 7396

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              | |          |||||||     |||||| ||||||||| || ||  | || || ||
Db       7397 GTGGAGCGTCACGGGGTGCGCCGGCTGCTGCTGGTGAGCCGCCGCGGGGCCGACGCCCCC 7456

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || |||| |  ||| | | ||| || || |   | || |||||||||  | | |  ||
Db       7457 GGCGCGGCCGACCTGGGCGAGGACCTCGCGGGCCTCGGCGCGGAGGTGGCGTTCGCCGCC 7516

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              || |  |  || ||   |   |||  |||  | | |     || || |||| | |
Db       7517 GCCGACGCCGCCGACCGCGAGAGCCTGGCGCGGGCGATCGCCACCGTGCCCGCCGAGCAT 7576

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |||||||||  |||||||  |||| || ||  | |||||  ||||||  | |    | |||
Db       7577 CCGCTGACGGCCGTCGTGCACACGGCGGGAGTCGTCGACGACGCGACGGTGGAGGCGCTC 7636

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              || || ||  |||||||  | ||||   |      | |||||  || |||||       ||| |
Db       7637 ACACCGGAACGGCTGGACGCGGTACTGCGCCCGAAGGTCGACGCCGCGTGGAACCTGCAC 7696

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||||  ||   ||   ||  |||     || |||||  || ||||||| | |  || |||
Db       7697 GAGCTCACCAAGGACCTGCGGCTCGACGCCTTCGTCCTCTTCTCCTCCGTCTCCGGCATC 7756

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | ||  | |||||  ||||  |||||||| ||||||||  | |   || |||||  || ||
Db       7757 GTCGGCACCGCCGGCCAGGCCAACTACGCGGCGGCCAACACGGGCCTCGACGCCCTCGCC 7816
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 17:56:34 ; Search time 63.1072 Seconds
(without alignments)
12253.192 Million cell updates/sec

Title: US-09-914-286-1_COPY_10609_11142
Perfect score: 534
Sequence: 1 gggtcggtgttggtgacggg.....cgctggcgtgggggctgtgg 534

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1029858 seqs, 724030393 residues

Total number of hits satisfying chosen parameters: 2059716

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	212.4	39.8	15872	9	US-09-860-846-1	Sequence 1, Appli
2	212.4	39.8	15872	9	US-09-988-384B-1	Sequence 1, Appli
3	212.4	39.8	15872	9	US-09-836-821-1	Sequence 1, Appli
4	212.4	39.8	15872	10	US-09-861-289-1	Sequence 1, Appli
5	207.4	38.8	4851	9	US-09-712-363-116	Sequence 116, App
6	199.6	37.4	13842	9	US-09-860-846-30	Sequence 30, Appl
7	199.6	37.4	13842	9	US-09-988-384B-30	Sequence 30, Appl
8	199.6	37.4	13842	9	US-09-836-821-30	Sequence 30, Appl
9	199.6	37.4	13842	10	US-09-861-289-30	Sequence 30, Appl
10	199.6	37.4	36778	9	US-09-860-846-5	Sequence 5, Appli
11	199.6	37.4	36778	9	US-09-836-821-5	Sequence 5, Appli
12	199.6	37.4	36778	10	US-09-861-289-5	Sequence 5, Appli
13	199.6	37.4	37948	9	US-09-988-384B-5	Sequence 5, Appli
14	199.6	37.4	38506	9	US-09-793-708-19	Sequence 19, Appl
15	196.4	36.8	11220	9	US-09-860-846-32	Sequence 32, Appl
16	196.4	36.8	11220	9	US-09-988-384B-32	Sequence 32, Appl
17	196.4	36.8	11220	9	US-09-836-821-32	Sequence 32, Appl
18	196.4	36.8	11220	10	US-09-861-289-32	Sequence 32, Appl
19	183.6	34.4	68750	9	US-10-014-717-1	Sequence 1, Appli
20	166	31.1	50937	9	US-09-808-880-1	Sequence 1, Appli
21	161.8	30.3	552	9	US-09-957-483-32	Sequence 32, Appl
22	157	29.4	552	9	US-09-957-483-34	Sequence 34, Appl
23	132	24.7	4689	9	US-09-860-846-34	Sequence 34, Appl
24	132	24.7	4689	9	US-09-988-384B-34	Sequence 34, Appl
25	132	24.7	4689	9	US-09-836-821-34	Sequence 34, Appl
26	132	24.7	4689	10	US-09-861-289-34	Sequence 34, Appl
27	87.6	16.4	12733	9	US-10-032-393-47	Sequence 47, Appl
28	87.6	16.4	12739	9	US-10-032-393-8	Sequence 8, Appli
c 29	83.8	15.7	1635	10	US-09-864-761-20241	Sequence 20241, A
c 30	83.8	15.7	1973	10	US-09-864-761-3471	Sequence 3471, Ap
31	80.8	15.1	5484	9	US-09-712-363-115	Sequence 115, App
32	79.8	14.9	1075	10	US-09-864-761-19241	Sequence 19241, A
33	79.8	14.9	1403	10	US-09-864-761-2513	Sequence 2513, Ap
34	74.4	13.9	446	10	US-09-864-761-20699	Sequence 20699, A
c 35	70.6	13.2	504	10	US-09-878-574-4330	Sequence 4330, Ap
c 36	66.8	12.5	152331	9	US-10-095-407-16	Sequence 16, Appl
37	66.6	12.5	830	10	US-09-864-761-19531	Sequence 19531, A
38	65.4	12.2	768	9	US-09-938-842A-812	Sequence 812, App
c 39	64.2	12.0	2283	9	US-09-938-842A-1691	Sequence 1691, Ap
c 40	64.2	12.0	2283	10	US-09-887-576-645	Sequence 645, App
c 41	62.6	11.7	4139	9	US-10-171-311-157	Sequence 157, App
c 42	62.6	11.7	4139	10	US-09-964-824A-105	Sequence 105, App
c 43	62.6	11.7	4139	10	US-09-964-824A-578	Sequence 578, App
c 44	62.6	11.7	4139	10	US-09-864-864-334	Sequence 334, App
c 45	62.6	11.7	4139	10	US-09-880-107-2121	Sequence 2121, Ap

ALIGNMENTS

RESULT 1
US-09-860-846-1
; Sequence 1, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-1

Query Match 39.8%; Score 212.4; DB 9; Length 15872;
Best Local Similarity 62.4%; Pred. No. 3.1e-42;
Matches 333; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

```
Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  | ||  | || || || || || || || |  || || |||    | || || || ||
Db       7337 GGCACCGTCCTCGTCACCGGGGGCACCGGAGGCCTCGGAGCGCTCCTCGCCCGCCACCTC 7396

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  |       |||||||    |||||| |||||||||| || ||  | || || ||
Db       7397 GTGGAGCGTCACGGGGTGCGCCGGCTGCTGCTGGTGAGCCGCCGCGGGGCCGACGCCCCC 7456

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || ||||  |  ||| | | ||| || || |   | || |||||||||  | | |  ||
Db       7457 GGCGCGGCCGACCTGGGCGAGGACCTCGCGGGCCTCGGCGCGGAGGTGGCGTTCGCCGCC 7516

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              || |  |  || ||   |  ||| ||| | | |     || || |||| |  |
Db       7517 GCCGACGCCGCCGACCGCGAGAGCCTGGCGCGGGCGATCGCCACCGTGCCCGCCGAGCAT 7576
```

THIS PAGE BLANK (USPTO)

```
;       TELEPHONE:  317-276-2459
;   INFORMATION FOR SEQ ID NO:  13:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  13987 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  350..13987
US-08-804-227C-13

  Query Match             37.1%;  Score 198;  DB 2;  Length 13987;
  Best Local Similarity   62.0%;  Pred. No. 2e-29;
  Matches  331;  Conservative    0;  Mismatches  200;  Indels    3;  Gaps    1;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  | ||  || | || || || || ||||  |||||||||  ||| || ||    ||
Db       7034 GGCACCGTACTGATCACCGGGGGCACCGGTGCCCTGGGTGCGCAGGTCGCCCGCAGGCTC 7093

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              ||  |   | | |  | |||   ||||||  ||||| |||||||| ||  |||  |  ||
Db       7094 GCCCGGTCGGGCGCCGCGCG---TCTGCTCCTGGTGGGCCGGCGCGGCGCGGCCGGCCCC 7150

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || | ||  |  ||    | ||||||| | ||| | ||  | || |||     | |  ||
Db       7151 GGAGTGGGCGAACTCGTCGAGGAGCTGACGGCGCTCGGTTCCGAAGTGGCCGTCGAGGCC 7210

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              || || || |  || |||   |   ||| |  ||| || | |||  | || |  | | |
Db       7211 TGCGACGTCGCCGACCGGGACGCACTGGCCGCGCTCCTCGCGGGCCTCCCCGAGGAGCGG 7270

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              || ||    |  ||  |||| ||||| |||||||| ||||| |    | ||| ||| |||
Db       7271 CCCCTCGTCGCCGTACTGCACGCGGCAGGTGTGCTCGACGACGGTGTGCTCGACTCGCTC 7330

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||  |||| ||| ||| | | ||  |   || ||||||    ||||| | |   |||||
Db       7331 ACCTCCGACCGGGTGGACGCCGTACTGCGGGACAAGGTCACCGCCGCCCGTCACCTGGAC 7390

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||||||||   ||  |  |||    || |||||||| ||||||||  ||  ||  |
Db       7391 GAGCTGACCGCGGACCTTCCGCTCGACGCCTTCGTGCTCTTCTCCTCCATCGTCGGCGTG 7450

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              |||    | ||||||| |  |||||| || ||||| ||||| || ||||| ||||||
Db       7451 TGGGGCAACGGAGGGCAGGCCGTCTACGCGGCCGCCAACGCCGCGCTCGACGCCCTGGCG 7510

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
               | ||||| |||||   ||.|    |  |  |  ||| | || ||||||| ||||
Db       7511 CAGCGGCGCCGGGCCAGGGGAGCCCGTGCCGCCTCGATCGCCTGGGGGCCGTGG 7564


Search completed: June 18, 2003, 01:08:14
Job time : 47.1323 secs
```

```
  Matches  335;  Conservative    0;  Mismatches  199;  Indels    6;  Gaps    1;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||  ||  ||||  ||  ||  ||  ||  |  |||||  |     |||  ||  |||||  ||
Db      14039 GGCACGGTCCTGCTGACCGGCGGCACCGGCGGCCTGGGCGGACTGGTCGCCCGGCACGTG 14098

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  ||  |  |||  ||  ||  ||     ||||||  ||||||||||||||  ||   ||||  ||  |||
Db      14099 GTGGGCGAGTGGGGCGTACGACGCCTGCTGCTGGTGAGCCGGCGGGGCACGGACGCCCCG 14158

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              ||  ||  ||  |   ||    |  |||||||   ||  ||||  ||  ||   |  |  |  |||
Db      14159 GGCGCCGACGAGCTCGTGCACGAGCTGGAGGCCCTGGGAGCCGACGTCTCGGTGGCCGCG 14218

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||  ||  ||  |   ||  ||     |    |    |     |  ||  ||  |    |  ||  ||||
Db      14219 TGCGACGTCGCCGACCGCGAAGCCCTCACCGCCGTACTCGACGCCATCCCCGCCGAACAC 14278

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |||||  ||  |   |||||  ||  ||||  ||  ||  ||     |||  |    ||  |||  |  |
Db      14279 CCGCTCACCGCGGTCGTCCACACGGCAGGCGTCCTCTCCGACGGCACCCTCCCGTCCATG 14338

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |||  |  |||     |||      ||   |   |  |||||||  ||  |||||    |   |  ||  ||
Db      14339 ACGACGGAGGACGTGGAACACGTACTGCGGCCCAAGGTCGACGCCGCGTTCCTCCTCGAC 14398

Qy        361 GAGCT------GACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCG 414
              ||  ||      |||||   |     ||  |||  |  ||||||||   ||||||||||  |||||
Db      14399 GAACTCACCTCGACGCCCGCATACGACCTGGCAGCGTTCGTCATGTTCTCCTCCGCCGCC 14458

Qy        415 GGGATCCTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCG 474
              |   ||  |  ||    ||  |||||||||   |||||||||  |||||  |  |  |||||
Db      14459 GCCGTCTTCGGTGGCGCGGGGCAGGGCGCCTACGCCGCCGCCAACGCCACCCTCGACGCC 14518

Qy        475 CTGGCGTACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||  ||  |  ||  ||  |||||  ||  ||  ||  ||  |   |  ||  ||  |  |||||  ||  |||
Db      14519 CTCGCCTGGCGCCGCCGGGCAGCCGGACTCCCCGCCCTCTCCCTCGGCTGGGGCCTCTGG 14578


RESULT 14
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
;  APPLICANT: ASHLEY, Gary
;  APPLICANT: BETLACH, Melanie C.
;  APPLICANT: BETLACH, Mary C.
;  APPLICANT: McDANIEL, Robert
;  APPLICANT: TANG, Li
;  TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 300622002120
;  CURRENT APPLICATION NUMBER: US/09/320,878A
;  CURRENT FILING DATE: 1999-05-27
;  EARLIER APPLICATION NUMBER: CIP OF 09/141,908
;  EARLIER FILING DATE: 1998-08-28
;  EARLIER APPLICATION NUMBER: CIP OF 09/073,538
;  EARLIER FILING DATE: 1998-05-06
;  EARLIER APPLICATION NUMBER: CIP OF 08/846,247
;  EARLIER FILING DATE: 1997-04-30
;  EARLIER APPLICATION NUMBER: 60/119,139
;  EARLIER FILING DATE: 1999-02-08
;  EARLIER APPLICATION NUMBER: 60/100,880
;  EARLIER FILING DATE: 1998-09-22
;  EARLIER APPLICATION NUMBER: 60/087,080
;  EARLIER FILING DATE: 1998-05-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
;   LENGTH: 38506
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-320-878-19


  Query Match           37.4%;  Score 199.6;  DB 3;  Length 38506;
  Best Local Similarity  62.0%;  Pred. No. 9.8e-30;
  Matches  335;  Conservative    0;  Mismatches  199;  Indels    6;  Gaps    1;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||  ||  ||||  ||  ||  ||  ||  |  |||||  |     |||  ||  |||||  ||
Db      12181 GGCACGGTCCTGCTGACCGGCGGCACCGGCGGCCTGGGCGGACTGGTCGCCCGGCACGTG 12240

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  ||  |  |||  ||  ||  ||     ||||||  ||||||||||||||  ||   ||||  ||  |||
Db      12241 GTGGGCGAGTGGGGCGTACGACGCCTGCTGCTGGTGAGCCGGCGGGGCACGGACGCCCCG 12300

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              ||  ||  ||  |   ||    |  |||||||   ||  ||||  ||  ||   |  |  |  |||
Db      12301 GGCGCCGACGAGCTCGTGCACGAGCTGGAGGCCCTGGGAGCCGACGTCTCGGTGGCCGCG 12360

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||  ||  ||  |   ||  ||     |    |    |     |  ||  ||  |    |  ||  ||||
Db      12361 TGCGACGTCGCCGACCGCGAAGCCCTCACCGCCGTACTCGACGCCATCCCCGCCGAACAC 12420

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |||||  ||  |   |||||  ||  ||||  ||  ||  ||     |||  |    ||  |||  |  |
Db      12421 CCGCTCACCGCGGTCGTCCACACGGCAGGCGTCCTCTCCGACGGCACCCTCCCGTCCATG 12480

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |||  |  |||     |||      ||   |   |  |||||||  ||  |||||    |   |  ||  ||
Db      12481 ACGACGGAGGACGTGGAACACGTACTGCGGCCCAAGGTCGACGCCGCGTTCCTCCTCGAC 12540

Qy        361 GAGCT------GACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCG 414
              ||  ||      |||||   |     ||  |||  |  ||||||||   ||||||||||  |||||
Db      12541 GAACTCACCTCGACGCCCGCATACGACCTGGCAGCGTTCGTCATGTTCTCCTCCGCCGCC 12600

Qy        415 GGGATCCTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCG 474
              |   ||  |  ||    ||  |||||||||   |||||||||  |||||  |  |  |||||
Db      12601 GCCGTCTTCGGTGGCGCGGGGCAGGGCGCCTACGCCGCCGCCAACGCCACCCTCGACGCC 12660

Qy        475 CTGGCGTACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||  ||  |  ||  ||  |||||  ||  ||  ||  ||  |   |  ||  ||  |  |||||  ||  |||
Db      12661 CTCGCCTGGCGCCGCCGGGCAGCCGGACTCCCCGCCCTCTCCCTCGGCTGGGGCCTCTGG 12720


RESULT 15
US-08-804-227C-13
; Sequence 13, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
```

```
US-09-105-537-1
; Sequence 1, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-1

  Query Match             39.8%;  Score 212.4;  DB 4;  Length 15872;
  Best Local Similarity   62.4%;  Pred. No. 4e-32;
  Matches  333;  Conservative    0;  Mismatches  201;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  | ||  | || || || || || || |  || || |||     | || || || ||
Db       7337 GGCACCGTCCTCGTCACCGGGGGCACCGGAGGCCTCGGAGCGCTCCTCGCCCGCCACCTC 7396

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  |         ||||||||    |||||| |||||||||| || ||  | || || ||
Db       7397 GTGGAGCGTCACGGGGTGCGCCGGCTGCTGCTGGTGAGCCGCCGCGGGGCCGACGCCCCC 7456

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || ||||  |  ||| | | ||| || || |   | || |||||||||  | | |  ||
Db       7457 GGCGCGGCCGACCTGGGCGAGGACCTCGCGGGCCTCGGCGCGGAGGTGGCGTTCGCCGCC 7516

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||  |  |  || ||    |    |||  |||  | | |     || || |||| |  |
Db       7517 GCCGACGCCGCCGACCGCGAGAGCCTGGCGCGGGCGATCGCCACCGTGCCCGCCGAGCAT 7576

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||||  ||||||||  |||| || ||  | ||||| ||||||  | |    | |||
Db       7577 CCGCTGACGGCCGTCGTGCACACGGCGGGAGTCGTCGACGACGCGACGGTGGAGGCGCTC 7636

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              || || || |||||||  | ||||  |     | |||||  || |||||       ||| |
Db       7637 ACACCGGAACGGCTGGACGCGGTACTGCGCCCGAAGGTCGACGCCGCGTGGAACCTGCAC 7696

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||||  ||   ||   ||  |||    || ||||| || ||||||||  | | |  || |||
Db       7697 GAGCTCACCAAGGACCTGCGGCTCGACGCCTTCGTCCTCTTCTCCTCCGTCTCCGGCATC 7756

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              |  ||  | ||||| |||| |||||||||  ||||||||  |  |  || |||||  || ||
Db       7757 GTCGGCACCGCCGGCCAGGCCAACTACGCGGCGGCCAACACGGGCCTCGACGCCCTCGCC 7816

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||  ||    ||  |||| ||| | |    |||||||| ||||| || |||
Db       7817 GCCCACCGCGCCGCCACGGGCCTGGCCGCCACGTCGCTGGCCTGGGGCCTCTGG 7870


RESULT 12
US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
;   LENGTH: 13842
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-30

  Query Match             37.4%;  Score 199.6;  DB 4;  Length 13842;
  Best Local Similarity   62.0%;  Pred. No. 9.8e-30;
  Matches  335;  Conservative    0;  Mismatches  199;  Indels    6;  Gaps    1;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||  || |||| || || || || |  ||||| |     ||| || |||||  ||
Db      12298 GGCACGGTCCTGCTGACCGGCGGCACCGGCGGCCTGGGCGGACTGGTCGCCCGGCACGTG 12357

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  || | ||| || || ||    |||||| ||||||||||||| ||  |||| || |||
Db      12358 GTGGGCGAGTGGGGCGTACGACGCCTGCTGCTGGTGAGCCGGCGGGGCACGGACGCCCCG 12417

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || || || |  ||    |   |||||||  ||  |||| || || ||   | | |  |||
Db      12418 GGCGCCGACGAGCTCGTGCACGAGCTGGAGGCCCTGGGAGCCGACGTCTCGGTGGCCGCG 12477

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              || || || |  || ||      |   |   |     | || || |   | || ||||
Db      12478 TGCGACGTCGCCGACCGCGAAGCCCTCACCGCCGTACTCGACGCCATCCCCGCCGAACAC 12537

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||| || |  ||||| ||  |||| || || ||   ||| |  ||  || | ||  |
Db      12538 CCGCTCACCGCGGTCGTCCACACGGCAGGCGTCCTCTCCGACGGCACCCTCCCGTCCATG 12597

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||| | |||     |||      ||  |   | |||||| || |||||   |  | || ||
Db      12598 ACGACGGAGGACGTGGAACACGTACTGCGGCCCAAGGTCGACGCCGCGTTCCTCCTCGAC 12657

Qy        361 GAGCT------GACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCG 414
              || ||      |||||  |      || ||| | |||||||  ||||||||| |||||
Db      12658 GAACTCACCTCGACGCCCGCATACGACCTGGCAGCGTTCGTCATGTTCTCCTCCGCCGCC 12717

Qy        415 GGGATCCTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCG 474
              |   || | ||    || |||||||||  |||||||| ||||| ||| | || |||||
Db      12718 GCCGTCTTCGGTGGCGCGGGGCAGGGCGCCTACGCCGCCGCCAACGCCACCCTCGACGCC 12777

Qy        475 CTGGCGTACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              || || |  || || ||||| || || || || |   | || || |  ||||| || |||
Db      12778 CTCGCCTGGCGCCGCCGGGCAGCCGGACTCCCCGCCCTCTCCCTCGGCTGGGGCCTCTGG 12837


RESULT 13
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-5

  Query Match             37.4%;  Score 199.6;  DB 4;  Length 36778;
  Best Local Similarity   62.0%;  Pred. No. 9.8e-30;
```

```
Qy        358 GATGAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGG 417
              | ||||||||||||||    |||| ||  ||||||||||||||||||| ||||| || ||
Db      19901 CACGAGCTGACGCGGGAGCTGGATCTCGCGGCGTTCGTGCTGTTCTCGTCGGCGGCCGGT 19960

Qy        418 ATCCTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTG 477
              |   | ||  |||| || ||| |||||||||| ||||| ||   |    || ||||||||
Db      19961 ACGTTTGGCGCGGCGGGCCAGAGCAACTACGCGGCGGCGAACACGTTCCTCGACGCGCTC 20020

Qy        478 GCGTACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              |||   |  ||||| |  || ||  ||| |  |  ||||||||||  | |||
Db      20021 GCGGCGCACCGGCGCGGCGGCGGGCTCGCGGCGACGAGCCTGGCGTGGGGCTTCTGG 20077


RESULT 9
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             41.0%;  Score 218.8;  DB 4;  Length 4403765;
  Best Local Similarity   63.1%;  Pred. No. 2.4e-33;
  Matches  337;  Conservative    0;  Mismatches  197;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  | ||| |||| || || ||||| ||  ||   ||| ||||||||||    ||| ||
Db    1871256 GGCACCGTGGTGGTCACTGGGGGGACCGGCATGGCCGGTTCGGCGGTGGCTACCCATTTG 1871315

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |   |       |||||||     |||||  |  ||||  |||||   | |||  | | ||
Db    1871316 GTCCGGCGACATGGGGTGGCCAATCTGGTTCTGGTCAGCCGAAGCGGTGAGCAGGCCGAC 1871375

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              | ||||  |   |  |||   | ||  || |  |||||  |||||  | | ||  |
Db    1871376 AGGGCGGCAGAAGTCGCGGCCCTGTTGCGCGAGGGCGGGGCCCAGGTGGCGGTGGTCTCC 1871435

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||  || ||    | | |||      || ||| |||| |   | |  | | |
Db    1871436 TGTGATGTGGCTGATCGTGATGCGCTGGCGGCATTGTTGGCGGGTCTGGATCCGCGCTAT 1871495

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |||||  |  || ||  | ||||| ||||| ||| ||||||||||    |||| |      |
Db    1871496 CCGCTTAAAGGGGTGTTTCATGCCGCTGGGGTGTTGGACGATGCCGTGATCACGGGCTTG 1871555

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              || || ||  ||| |||  |||||||| |||||||| ||||  ||       ||| |
Db    1871556 ACACCGGATCGGGTGGATACGGTGTTGCGGGCCAAGGTCGATGGGGCCTGGAATCTGCAC 1871615

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||||  ||   ||  |||||  |||||||||| ||| |||| || |||   || |||||
Db    1871616 GAGCTAACCGAGGACATGGATTTGTCGGCGTTTGTGGTGTTTTCGTCGATGGCCGGGATT 1871675

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              |||| |  | |  ||||| || || || |||||  |||||   | ||||||  | ||| |
Db    1871676 GTGGGCACACCGGCTCAGGGGAATTATGCTGCGGCGAATGCGTTTTTGGACGGGTTGGTG 1871735

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              |    |||||  ||   || ||| | ||  ||||| |||||||||| ||||||
Db    1871736 GCCTATCGGCGCTCGCGTGGGCTGGCCGGATTGTCGGTGGCGTGGGGACTGTGG 1871789


RESULT 10
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             41.0%;  Score 218.8;  DB 4;  Length 4411529;
  Best Local Similarity   63.1%;  Pred. No. 2.4e-33;
  Matches  337;  Conservative    0;  Mismatches  197;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  | ||| |||| || || ||||| ||  ||   ||| ||||||||||    ||| ||
Db    1880381 GGCACCGTGGTGGTCACTGGGGGGACCGGCATGGCCGGTTCGGCGGTGGCTACCCATTTG 1880440

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |   |       |||||||     |||||  |  ||||  |||||   | |||  | | ||
Db    1880441 GTCCGGCGACATGGGGTGGCCAATCTGGTTCTGGTCAGCCGAAGCGGTGAGCAGGCCGAC 1880500

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              | ||||  |   |  |||   | ||  || |  |||||  |||||  | | ||  |
Db    1880501 AGGGCGGCAGAAGTCGCGGCCCTGTTGCGCGAGGGCGGGGCCCAGGTGGCGGTGGTCTCC 1880560

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||  || ||    | | |||      || ||| |||| |   | |  | | |
Db    1880561 TGTGATGTGGCTGATCGTGATGCGCTGGCGGCATTGTTGGCGGGTCTGGATCCGCGCTAT 1880620

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |||||  |  || ||  | ||||| ||||| ||| ||||||||||    |||| |      |
Db    1880621 CCGCTTAAAGGGGTGTTTCATGCCGCTGGGGTGTTGGACGATGCCGTGATCACGGGCTTG 1880680

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              || || ||  ||| |||  |||||||| |||||||| ||||  ||       ||| |
Db    1880681 ACACCGGATCGGGTGGATACGGTGTTGCGGGCCAAGGTCGATGGGGCCTGGAATCTGCAC 1880740

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||||  ||   ||  |||||  |||||||||| ||| |||| || |||   || |||||
Db    1880741 GAGCTAACCGAGGACATGGATTTGTCGGCGTTTGTGGTGTTTTCGTCGATGGCCGGGATT 1880800

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              |||| |  | |  ||||| || || || |||||  |||||   | ||||||  | ||| |
Db    1880801 GTGGGCACACCGGCTCAGGGGAATTATGCTGCGGCGAATGCGTTTTTGGACGGGTTGGTG 1880860

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              |    |||||  ||   || ||| | ||  ||||| |||||||||| ||||||
Db    1880861 GCCTATCGGCGCTCGCGTGGGCTGGCCGGATTGTCGGTGGCGTGGGGACTGTGG 1880914


RESULT 11
```

```
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-198-1

  Query Match             43.1%;  Score 230;  DB 2;  Length 44377;
  Best Local Similarity   64.4%;  Pred. No. 2e-35;
  Matches  344;  Conservative    0;  Mismatches  190;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  | ||| || | || || || || || | ||||  ||||      || ||||| |||
Db      18618 GGCACCGTGCTGATCACCGGCGGCACCGGAGCACTGGCCGCGGAAACCGCCCGGCACCTG 18677

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  |     |      ||||||  |||| ||| ||||| || |||| ||||| || || ||
Db      18678 GTCGACCGGCACAAGGTGCGCCATCTCCTGCTGGTGGGCAGGCGCGGTCCCGACGCACCC 18737

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || |  ||  | |||   || ||| || ||| || ||| || |||||     |    ||
Db      18738 GGCGTCGATCGACTGGTCGCCGAGTTGACCGAGTCGGGTGCCGAGGTCGCCGTACGGGCC 18797

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              |||||  ||   ||| ||    |   ||  ||| ||||| || |   | || | ||
Db      18798 TGTGACGTCACGGACCGCGACGCCCTGCGCCGCCTGCTCGACGCACTCCCCGACGAACAC 18857

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||  | ||  |||||  | ||  ||  |||||  |||||  |    | ||  || | |
Db      18858 CCGCTGACCTGCGTGGTGCACACCGCCGGGGTGCTCGACGACGGCGTGCTCTCCGCCCAG 18917

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |||  ||||||||  | |  |||||||  ||   |  ||||||   ||  |||||   |     |||||
Db      18918 ACGGCCGAGCGGATCGACACGGTGCTCCGGCCCAAGGCCGACGCCGCCGTCCACCTGGAC 18977

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||||||||  ||||   ||  |   |  |   |  |   |  |||||||  ||||||||  |  |||   | |
Db      18978 GAGCTGACCCGGGAGATCGGACGGGTGCCCCTGGTGCTGTACTCCTCGGTCTCGGCCACC 19037

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              |||||     ||  |||||||  |    |||||  ||||||||  |||     ||||||||||||
Db      19038 CTGGGCAGCGCGGGGCAGGCCGGGTACGCGGCGGCCAACGCCTTCATGGACGCGCTGGCC 19097

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              |||||||  | ||  ||  ||  |   ||  | |  |||||||  |   |||||    ||||
Db      19098 GCCCGGCGGTGCGCCGCCGGGCACCCCGCGCTGTCGCTCGGCTGGGGCTGGTGG 19151


RESULT 8
US-09-144-085-3
; Sequence 3, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
;  APPLICANT: Gustafsson, Claes
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Ashley, Gary
;  APPLICANT: Julien, Bryan
;  APPLICANT: Ziermann, Rainer
;  TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: 30062-20020.20
;  CURRENT APPLICATION NUMBER: US/09/144,085
;  CURRENT FILING DATE: 1998-08-31
;  EARLIER APPLICATION NUMBER: 09/010,809
;  EARLIER FILING DATE: 1998-01-22
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 33529
;   TYPE: DNA
;   ORGANISM: Sorangium cellulosum
US-09-144-085-3

  Query Match             41.2%;  Score 220.2;  DB 4;  Length 33529;
  Best Local Similarity   64.4%;  Pred. No. 1.4e-33;
  Matches  346;  Conservative    0;  Mismatches  188;  Indels    3;  Gaps    1;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||| |||||||  || |||||  || | |||||||  ||||  | |||  ||| |||
Db      19541 GGCACGGTGCTGGTGACCGGCGGGACAGGGGAGCTGGGTCAGGCGATCGCGGCGCACCTG 19600

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |   | | |    ||||||  || |  ||   || ||   |    || || || | ||| || ||
Db      19601 GTGCGCGCGCACGGGGTGAGGCACCTTGTGCTGACGTCGCGTCGCGGGCTGGAGGCGCCC 19660

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGA---GGTGCGGATTGTT 177
              || ||   |  ||  |  |   || |  | | | || ||  ||||  | | | |
Db      19661 GGGGCCCGCGAGCTCGTGCAGTCACTCGAGGAGCTCGGCGCCGAGACGGTGACGGTGGCT 19720

Qy        178 GCGTGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGG 237
              ||||| || |||  | |||||  ||||||  |  ||  ||||||  ||  |    || | |
Db      19721 GCGTGCGACGTGTCGAAGCGGGAGGAGGTCGCGCGTGTGCTGGCCGGCATCGACGCGGCG 19780

Qy        238 TGTCCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCT 297
              |||||||||  |  ||  ||||   ||| ||.|||| ||||| |     | || || |
Db      19781 CGTCCGCTGAGCGCGGTGCTGCACCTGGCCGGCGTGCTCGACGACGGCGTGCTCACCGCC 19840

Qy        298 CTCACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTG 357
              |  ||| |||||||  ||       |||| | ||| | ||||||||  |. || ||   |||
Db      19841 CAGACGGCCGAGCGCCTCTCGCGGGTGCTGGCGCCGAAGGTGGACGGGGCGCTGCACCTG 19900
```

```
US-09-370-700-1

  Query Match             44.9%;  Score 239.6;  DB 4;  Length 80161;
  Best Local Similarity   65.5%;  Pred. No. 3.2e-37;
  Matches  350;  Conservative    0;  Mismatches  184;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  | ||| |||| || || || || || | ||| || ||   | | || || ||||||
Db      74599 GGCACCGTGCTGGTCACCGGCGGAACCGGCGGGCTCGGGGCCCTGCTCGCACGCCATCTG 74658

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  | |      || |||||| | ||| || | |  || || ||||||| ||  || ||
Db      74659 GTCAGCGCCCACGGTGTGCGGCACCTGTTGCTCGCAAGTCGCCGTGGTCTGGCGGCGCCC 74718

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || ||||| |  |||  ||| |||||||    |   || || || ||   | | || |||
Db      74719 GGAGCGGATGAGCTGGTGGCCGAGCTGGAGCAGGCCGGCGCCGACGTCGCGGTCGTCGCG 74778

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              || ||   ||  || |||      | | |  |||||| ||| |   || |||||
Db      74779 TGCGACTCGGCAGATCGGGACTCGCTTGCGCGGCTGGTGGCGTCGGTGCCTGCGGAAAAC 74838

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||| ||  ||  || ||||| || || ||||||||||| || |     | |    ||  |
Db      74839 CCGTTGCGGGTGGTGGTGCACGCCGCCGGTGTGCTGGATGACGGTGTGCTGATGTCGATG 74898

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||||  |||||  ||| | |||||||   | |||| ||||||||||    |  ||| |
Db      74899 TCGCCGGAGCGCTTGGACGCGGTGTTGCGGCCCAAAGTGGATGCCGCGTGGTACCTGCAC 74958

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||||||| ||||   | |  ||||||||||||||| ||||||||||| ||||||  |
Db      74959 GAGCTGACTCGGGAACTCGGTCTGTCGGCGTTCGTGTTGTTCTCCTCGGTCGCGGGCCTG 75018

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | ||    || |||||| |||| ||||| || | ||| ||    |||| ||  |||||
Db      75019 TTCGGCGGTGCGGGGCAGAGCAATTACGCTGCCGGCAACGCTTTCCTGGATGCCTTGGCG 75078

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              | | |||| |||   ||| ||||| | | |||||||||| |  |||||||||
Db      75079 CATTGCCGGCAGGCCCAGGGGCTGCCCGCGCTGTCGCTGGCCTCCGGGCTGTGG 75132


RESULT 6
US-08-804-227C-7
; Sequence 7, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44377 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  350..14002
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14046..20036
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20110..31284
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  31329..36071
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  36155..41830
US-08-804-227C-7

  Query Match             43.1%;  Score 230;  DB 2;  Length 44377;
  Best Local Similarity   64.4%;  Pred. No. 2e-35;
  Matches  344;  Conservative    0;  Mismatches  190;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  | ||| || | || || || || || |  ||||  ||||      || ||||| |||
Db      18618 GGCACCGTGCTGATCACCGGCGGCACCGGAGCACTGGCCGCGGAAACCGCCCGGCACCTG 18677

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  |    |    ||||||  |||| ||| ||||| || |||| ||||| || || ||
Db      18678 GTCGACCGGCACAAGGTGCGCCATCTCCTGCTGGTGGGCAGGCGCGGTCCCGACGCACCC 18737

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || |  ||  | |||    || ||| || ||| || ||| || |||||      |    ||
Db      18738 GGCGTCGATCGACTGGTCGCCGAGTTGACCGAGTCGGGTGCCGAGGTCGCCGTACGGGCC 18797

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              |||||  ||    |||  ||      |   ||  |||  |||| || |    | || | ||
Db      18798 TGTGACGTCACGGACCGCGACGCCCTGCGCCGCCTGCTCGACGCACTCCCCGACGAACAC 18857

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||  | ||  |||||  | ||  ||  |||||  |||||  |    |  || ||  | |
Db      18858 CCGCTGACCTGCGTGGTGCACACCGCCGGGGTGCTCGACGACGGCGTGCTCTCCGCCCAG 18917

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||| |||||||| | |  |||||||  ||   |  ||||||   ||  |||||   |     |||||
Db      18918 ACGGCCGAGCGGATCGACACGGTGCTCCGGCCCAAGGCCGACGCCGCCGTCCACCTGGAC 18977

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||||||||  ||||   ||  |  |  |  |  | |||||||  ||||||||  | |||   | |
Db      18978 GAGCTGACCCGGGAGATCGGACGGGTGCCCCTGGTGCTGTACTCCTCGGTCTCGGCCACC 19037

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              |||||    || |||||||  |   |||||  ||||||||  |||    ||||||||||||
Db      19038 CTGGGCAGCGCGGGGCAGGCCGGGTACGCGGCGGCCAACGCCTTCATGGACGCGCTGGCC 19097

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              |||||||  | ||  ||  ||  |   ||  | | |||||||  |  |||||   ||||
Db      19098 GCCCGGCGGTGCGCCGCCGGGCACCCCGCGCTGTCGCTCGGCTGGGGCTGGTGG 19151


RESULT 7
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
```

```
Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  | |     |||||||||  |||||  ||||||||||||||||  || ||||||||||
Db    2298097 GTGGCTCGTCATGGGGTGCGCAATCTGGTGTTGGTGAGCCGGCGCGGCCCGGATGCTCCC 2298038

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || || | ||  |||  ||| ||| |||||||     || ||| |||| | | | || ||
Db    2298037 GGGGCTGCGGAGCTGGTGGCCGAGTTGGCCGCCGCCGGTGCGCAGGTACAGGTGGTGGCC 2297978

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||| | || ||| ||   || | ||| |  | || | |  | | |||| |    |  |
Db    2297977 TGTGACGCGGCGGATCGAGCGGCGTTGGCCAAGGTGATCGCCGATATTCCGGTTCAGCAT 2297918

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||  || |||| ||  | ||  | || || |  || ||||| ||| | |    || ||
Db    2297917 CCATTGTCGGGCGTGATCCACACCGCCGGCGCACTCGACGACGCGGTGGTGATGTCACTG 2297858

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              || || || ||| |||    ||||||    | ||||||||| || ||        ||| |
Db    2297857 ACACCGGATCGGGTGGATGTGGTGTTGCGGTCCAAGGTGGACGCGGCGTGGCACCTGCAC 2297798

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||| |||| || |  ||||  |||||||||| ||  |||| || |||   || ||| |
Db    2297797 GAGTTGACTCGCGACCTGGATGTGTCGGCGTTTGTCATGTTTTCGTCGATGGCCGGGCTG 2297738

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | || ||| | || |||| |||||| || || ||||||  |   | |||| ||||||||
Db    2297737 GTCGGATCGTCGGGCCAGGCCAACTATGCGGCCGCCAATTCGTTTTTGGATGCGCTGGCC 2297678

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                ||  ||||||||     || |||||||    | || ||||  ||||| ||||||
Db    2297677 GCCCACCGGCGGGCCCATGGGCTGCCGGCCATCTCCCTGGGCTGGGGTCTGTGG 2297624


RESULT 4
US-09-036-987A-1
; Sequence 1, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  80161 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-036-987A-1

  Query Match             44.9%;  Score 239.6;  DB 3;  Length 80161;
  Best Local Similarity   65.5%;  Pred. No. 3.2e-37;
  Matches  350;  Conservative    0;  Mismatches  184;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  | ||| |||| || || || || || | ||| || ||    | | || || ||||||
Db      74599 GGCACCGTGCTGGTCACCGGCGGAACCGGCGGGCTCGGGGCCCTGCTCGCACGCCATCTG 74658

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |   | |       || |||||| | ||| || | |  || || ||||||| ||  || ||
Db      74659 GTCAGCGCCCACGGTGTGCGGCACCTGTTGCTCGCAAGTCGCCGTGGTCTGGCGGCGCCC 74718

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || ||||| |  |||  ||| |||||||     |   || || || ||   | | || |||
Db      74719 GGAGCGGATGAGCTGGTGGCCGAGCTGGAGCAGGCCGGCGCCGACGTCGCGGTCGTCGCG 74778

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              || ||    ||  || |||      | | |  |||||| ||| |   || ||||| |
Db      74779 TGCGACTCGGCAGATCGGGACTCGCTTGCGCGGCTGGTGGCGTCGGTGCCTGCGGAAAAC 74838

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||| ||  ||  || ||||| || || |||||||||| || |    | |    || |
Db      74839 CCGTTGCGGGTGGTGGTGCACGCCGCCGGTGTGCTGGATGACGGTGTGCTGATGTCGATG 74898

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |||| |||||  ||| | |||||||   | |||| ||||||||||     |  ||| |
Db      74899 TCGCCGGAGCGCTTGGACGCGGTGTTGCGGCCCAAAGTGGATGCCGCGTGGTACCTGCAC 74958

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||||||| ||||   | |  |||||||||||||||| |||||||||| ||||||  |
Db      74959 GAGCTGACTCGGGAACTCGGTCTGTCGGCGTTCGTGTTGTTCTCCTCGGTCGCGGGCCTG 75018

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | ||    || |||||| |||| ||||| || | ||| ||    ||||| ||  |||||
Db      75019 TTCGGCGGTGCGGGGCAGAGCAATTACGCTGCCGGCAACGCTTTCCTGGATGCCTTGGCG 75078

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              |  | |||| |||   ||| ||||| | | ||||||||| |  |||||||||
Db      75079 CATTGCCGGCAGGCCCAGGGGCTGCCCGCGCTGTCGCTGGCCTCCGGGCTGTGG 75132


RESULT 5
US-09-370-700-1
; Sequence 1, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 80161
;   TYPE: DNA
;   ORGANISM: Saccharopolyspora spinosa
```

```
;       NAME/KEY:  CDS
;       LOCATION:  31232..36067
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  36249..41774
US-08-804-227C-1

  Query Match             47.0%;  Score 250.8;  DB 2;  Length 43280;
  Best Local Similarity   66.9%;  Pred. No. 2.6e-39;
  Matches  357;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||  |||||||||| | ||| ||   ||| ||   |  | |||| |||||||||
Db      12864 GGCACGGTCCTGGTGACGGGCGCGACCGGCACGCTCGGGCGGCTGCTGGCCCGGCATCTG 12923

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              || |  | |   || |||||  |||||||| || |||||||  | || | || |||
Db      12924 GCCGCCGAGCACGGTGTGCGCCATCTGCTGCTGCTGAGCCGCGGCGGCCGGGCTGCCGAA 12983

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || ||||| |  ||     ||||| |||||||| |||||  || |||  | |  | |  ||
Db      12984 GGCGCGGACGAACTCGCCGCGGAACTGGCCGGGTTGGAAGCCGAGCCGTGCTTCGCGGCC 13043

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              |||||  | || |||  ||    ||   |||  ||| |||||| ||  || || ||||   |
Db      13044 TGTGACGCGGCGGACCGCGAGGCCCTGGCACGGGTGCTGGCGGAGGTGCCGGCCGACCGG 13103

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||  || ||   | || ||||| || ||||| |||||  |   ||  ||| | | ||
Db      13104 CCGCTGACCGGAGTGATCCACGCGGCCGGGGTGCTCGACGACGGCACACTCGACGCGCTG 13163

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              || || || ||| | || || ||  |   | | |||| ||| || || ||     ||| |
Db      13164 ACCCCGGAACGGATCGGTACCGTCATGCGGCCGAAGGCGGACGCGGCGCTGAACCTGCAC 13223

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              || ||||| |||   |   |||||||| ||||| | | |||||    || || || |||
Db      13224 GAACTGACCCGGACCAGCCCGCTGTCGGTGTTCGCGGTCTTCTCGGGCGCTGCCGGCATC 13283

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              |||||     |||| |||| ||||||||||||| |||||  ||   || ||||||| |||
Db      13284 CTGGGCCGCCCCGGACAGGCCAACTACGCCGCCGCCAACACCTTCCTCGACGCGCTCGCG 13343

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                | |  || || ||    || || || |  |||||||||| ||||||||||||
Db      13344 CAGCACCGCCGCGCCCACGGCCTCCCCGCCGTGTCGCTGGCCTGGGGGCTGTGG 13397


RESULT 2
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match             47.0%;  Score 250.8;  DB 4;  Length 4411529;
  Best Local Similarity   66.9%;  Pred. No. 2.5e-39;
  Matches  357;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||  |||||||||| | ||| ||   ||| ||   |  | |||| |||||||||
Db    2295857 GGCACGGTCCTGGTGACGGGCGCGACCGGCACGCTCGGGCGGCTGCTGGCCCGGCATCTG 2295798

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              || |  | |   || |||||  |||||||| || |||||||  | || | || |||
Db    2295797 GTGGCTCGTCATGGGGTGCGCAATCTGGTGTTGGTGAGCCGGCGCGGCCCGGATGCTCCC 2295738

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || ||  | ||  |||  ||| ||| |||||||     || ||| |||||| | | || ||
Db    2295737 GGGGCTGCGGAGCTGGTGGCCGAGTTGGCCGCCGCCGGTGCGCAGGTGCAGGTGGTGGCC 2295678

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              |||||  | || |||  ||    ||   |||  ||| |||||| ||  || || ||||   |
Db    2295677 TGTGACGCGGCGGACCGCGAGGCGGCGTTGGCCAAGGTGATCGCCGATATTCCGGTTCAGCAT 2295618

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||  || |||| ||   | ||  | || || |  || ||||| |||  | |    || ||
Db    2295617 CCATTGTCGGGCGTGATCCACACCGCCGGCGCACTCGACGACGCGGTGGTGATGTCACTG 2295558

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              || || || ||| |||     ||||||   | |||||||||| || ||        ||| |
Db    2295557 ACACCGGATCGGGTGGATGTGGTGTTGCGGTCCAAGGTGGACGCGGCGTGGCACCTGCAC 2295498

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||| |||| || |   |||  |||||||||| ||  |||| || |||   || ||| |
Db    2295497 GAGTTGACTCGCGACCTGGATGTGTCGGCGTTTGTCATGTTTTCGTCGATGGCCGGGCTG 2295438

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | || ||| | || |||| |||||| || || |||||  |   | |||| |||||||||
Db    2295437 GTCGGATCGTCGGGCCAGGCCAACTATGCGGCCGCCAATTCGTTTTTGGATGCGCTGGCC 2295378

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                ||  |||||||||     || |||||||   | || ||||  |||||  ||||||
Db    2295377 GCCCACCGGCGGGCCCATGGGCTGCCGGCCATCTCCCTGGGCTGGGGTCTGTGG 2295324


RESULT 3
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match             46.7%;  Score 249.2;  DB 4;  Length 4403765;
  Best Local Similarity   66.7%;  Pred. No. 5.1e-39;
  Matches  356;  Conservative    0;  Mismatches  178;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||  |||||||||| | ||| ||   ||| ||   |  | |||| |||||||||
Db    2298157 GGCACGGTCCTGGTGACGGGCGCGACCGGCACGCTCGGGCGGCTGCTGGCCCGGCATCTG 2298098
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:45:13 ; Search time 23.1323 Seconds
(without alignments)
7079.508 Million cell updates/sec

Title: US-09-914-286-1_COPY_10609_11142
Perfect score: 534
Sequence: 1 gggtcggtgttggtgacggg.....cgctggcgtgggggctgtgg 534

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	250.8	47.0	43280	2	US-08-804-227C-1	Sequence 1, Appli
c 2	250.8	47.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c 3	249.2	46.7	4403765	4	US-09-103-840A-2	Sequence 2, Appli
4	239.6	44.9	80161	3	US-09-036-987A-1	Sequence 1, Appli
5	239.6	44.9	80161	4	US-09-370-700-1	Sequence 1, Appli
6	230	43.1	44377	2	US-08-804-227C-7	Sequence 7, Appli
7	230	43.1	44377	2	US-08-804-198-1	Sequence 1, Appli
8	220.2	41.2	33529	4	US-09-144-085-3	Sequence 3, Appli
9	218.8	41.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
10	218.8	41.0	4411529	4	US-09-103-840A-1	Sequence 1, Appli
11	212.4	39.8	15872	4	US-09-105-537-1	Sequence 1, Appli
12	199.6	37.4	13842	4	US-09-105-537-30	Sequence 30, Appl
13	199.6	37.4	36778	4	US-09-105-537-5	Sequence 5, Appli
14	199.6	37.4	38506	3	US-09-320-878-19	Sequence 19, Appl
15	198	37.1	13987	2	US-08-804-227C-13	Sequence 13, Appl
16	196.4	36.8	11220	4	US-09-105-537-32	Sequence 32, Appl
17	192	36.0	20235	1	US-07-642-734C-3	Sequence 3, Appli
18	192	36.0	20235	3	US-08-439-009A-3	Sequence 3, Appli
19	183.6	34.4	68750	3	US-09-335-409-1	Sequence 1, Appli
20	183.6	34.4	68750	4	US-09-568-102-1	Sequence 1, Appli
21	183.6	34.4	68750	4	US-09-567-969-1	Sequence 1, Appli
22	183.6	34.4	68750	4	US-09-568-480-1	Sequence 1, Appli
23	183.6	34.4	68750	4	US-09-568-486-1	Sequence 1, Appli
24	183.6	34.4	68750	4	US-09-568-472-1	Sequence 1, Appli
25	183.6	34.4	68750	4	US-09-567-899-1	Sequence 1, Appli
26	182	34.1	71989	4	US-09-443-501A-2	Sequence 2, Appli
27	178.8	33.5	28958	1	US-08-258-261B-6	Sequence 6, Appli
28	178.8	33.5	28958	1	US-08-456-837-6	Sequence 6, Appli
29	178.8	33.5	28958	1	US-08-457-342-6	Sequence 6, Appli
30	178.8	33.5	28958	1	US-08-457-646A-6	Sequence 6, Appli
31	178.8	33.5	28958	1	US-08-458-076A-6	Sequence 6, Appli
32	178.8	33.5	28958	1	US-08-764-233A-4	Sequence 4, Appli
33	178.8	33.5	28958	1	US-08-457-335A-6	Sequence 6, Appli
34	178.8	33.5	28958	1	US-08-729-214-6	Sequence 6, Appli
35	178.8	33.5	28958	3	US-09-028-934-6	Sequence 6, Appli
36	178.8	33.5	49377	1	US-08-764-233A-1	Sequence 1, Appli
37	174	32.6	11219	1	US-07-642-734C-1	Sequence 1, Appli
38	174	32.6	11219	3	US-08-439-009A-1	Sequence 1, Appli
39	166	31.1	50937	4	US-09-428-517-1	Sequence 1, Appli
40	158.8	29.7	477	3	US-09-010-809-2	Sequence 2, Appli
c 41	158.8	29.7	477	3	US-09-010-809-17	Sequence 17, Appl
42	132	24.7	4689	4	US-09-105-537-34	Sequence 34, Appl
43	90.4	16.9	897	4	US-09-434-288-6	Sequence 6, Appli
44	81.4	15.2	1681	4	US-09-434-288-7	Sequence 7, Appli
45	72	13.5	320	4	US-09-165-264-13	Sequence 13, Appl

ALIGNMENTS

```
RESULT 1
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  43280 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  816..14234
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  14351..19945
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  20010..31199
;    FEATURE:
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:44 ; Search time 1035.85 Seconds
(without alignments)
15003.005 Million cell updates/sec

Title: US-09-914-286-1_COPY_10609_11142
Perfect score: 534
Sequence: 1 gggtcggtgttggtgacggg.....cgctggcgtgggggctgtgg 534

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		534	100.0	12381	6	AX006889	AX006889 Sequence
2		534	100.0	30690	6	E38020	E38020 Avermectin
3		534	100.0	64957	1	AB032367	AB032367 Streptomy
4		532.4	99.7	31422	6	E38021	E38021 Avermectin
5	c	532.4	99.7	64957	1	AB032367	AB032367 Streptomy
6	c	258.8	48.5	39314	1	SGR300302	AJ300302 Streptomy
7		257.2	48.2	2700	6	AX211733	AX211733 Sequence
8		257.2	48.2	2706	1	AF420594	AF420594 Streptomy
9		257.2	48.2	65140	6	AX211705	AX211705 Sequence
10		257.2	48.2	123580	1	AF263912	AF263912 Streptomy
11		257.2	48.2	125401	6	AX211739	AX211739 Sequence
12		252	47.2	113193	1	AF357202	AF357202 Streptomy
13		250.8	47.0	43280	1	SFU78289	U78289 Streptomyce
14	c	250.8	47.0	53450	1	MTV018	AL021899 Mycobacte
15	c	249.2	46.7	15970	1	AE007061	AE007061 Mycobacte
16		244	45.7	27541	6	AX211706	AX211706 Sequence
17		239.6	44.9	16767	6	AX089464	AX089464 Sequence
18		239.6	44.9	29736	6	AX089421	AX089421 Sequence
19		239.6	44.9	50000	6	AX089417	AX089417 Sequence
20		239.6	44.9	80161	1	AY007564	AY007564 Saccharop
21		239.6	44.9	80161	6	AR165018	AR165018 Sequence
22	c	236.4	44.3	26195	1	SC1G7	AL591083 Streptomy
23		234.8	44.0	30000	6	AX250262	AX250262 Sequence
24		234.8	44.0	41097	1	AF016585	AF016585 Streptomy
25		230.8	43.2	3225	1	AF232752	AF232752 Mycobacte
26	c	230.8	43.2	104326	1	AB070940	AB070940 Streptomy
27		230	43.1	14775	6	AX089462	AX089462 Sequence
28		230	43.1	50000	6	AX089420	AX089420 Sequence
29	c	228.4	42.8	24225	1	SC2C4	AL512902 Streptomy
30		227	42.5	35445	1	MLCB2052	Z98604 Mycobacteri
31	c	227	42.5	318200	1	MLEPRTN6	AL583922 Mycobacte
32		225.2	42.2	30000	6	AX250263	AX250263 Sequence
33		225.2	42.2	41782	1	AY118081	AY118081 Streptomy
34		223.6	41.9	78210	1	AB070949	AB070949 Streptomy
35	c	223.6	41.9	84985	1	SNA278573	AJ278573 Streptomy
36		223.2	41.8	20394	1	SNA132222	AJ132222 Streptomy
37		223.2	41.8	20394	6	AX067996	AX067996 Sequence
38		220.2	41.2	33529	6	AR166425	AR166425 Sequence
39		218.8	41.0	18614	1	AE007033	AE007033 Mycobacte
40		218.8	41.0	38000	1	MTCY06H11	Z85982 Mycobacteri
41		218.8	41.0	104326	1	AB070940	AB070940 Streptomy
42		217.2	40.7	1784	1	AF411573S3	AF411575 Actinomad
43		217.2	40.7	25150	1	AB070945	AB070945 Streptomy
44		215.8	40.4	17101	1	AB017641	AB017641 Micromono
45	c	208	39.0	82746	1	AF453501	AF453501 Actinosyn

ALIGNMENTS

RESULT 1
AX006889
LOCUS AX006889 12381 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 1 from Patent WO0001827.
ACCESSION AX006889
VERSION AX006889.1 GI:9994904
KEYWORDS .
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 12381)
AUTHORS Kellenberger,J.L., Leadlay,P.F., Staunton,J., Mcarthur,H.A. and
Stutzman-Engwall,K.J.
TITLE Polyketides, their preparation, and materials for use therein

```
  JOURNAL   Patent: WO 0001827-A 1 13-JAN-2000;
            KELLENBERGER JOHANNES LAURENZ (CH); LEADLAY PETER FRANCIS (GB);
            STAUNTON JAMES (GB); BIOTICA TECH LTD (GB); PFIZER (US); MCARTHUR
            HAMISH ALASTAIR IRVIN (US); STUTZMAN ENGWALL KIM JONELLE (US)
FEATURES             Location/Qualifiers
     source          1. .12381
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     1884 a   4561 c   4005 g   1931 t
ORIGIN

  Query Match             100.0%;  Score 534;  DB 6;  Length 12381;
  Best Local Similarity   100.0%;  Pred. No. 5.1e-69;
  Matches  534;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11426 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 11485

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11486 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 11545

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11546 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 11605

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11606 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 11665

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11666 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 11725

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11726 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 11785

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11786 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 11845

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11846 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 11905

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11906 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 11959


RESULT 2
E38020
LOCUS       E38020                30690 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Avermectin aglycon synthase gene.
ACCESSION   E38020
VERSION     E38020.1  GI:18626909
KEYWORDS    JP 2000245457-A/1.
SOURCE      Streptomyces avermitilis.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 30690)
  AUTHORS   Omura,S. and Ikeda,H.
  TITLE     Avermectin aglycon synthase gene
  JOURNAL   Patent: JP 2000245457-A 1 12-SEP-2000;
            THE KITASATO INSTITUTE
COMMENT     OS   Streptomyces avermitilis
            PN   JP 2000245457-A/1
            PD   12-SEP-2000
            PF   24-FEB-1999 JP 1999046961
            PR
            PI   SATOSHI OMURA,HARUO IKEDA
            PC   C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
            C12Q1/68//C07D493/22,
            PC   (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(11916)
            FT   CDS             (11971). .(30687).
FEATURES             Location/Qualifiers
     source          1. .30690
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
BASE COUNT     5355 a  12455 c   8617 g   4263 t
ORIGIN

  Query Match             100.0%;  Score 534;  DB 6;  Length 30690;
  Best Local Similarity   100.0%;  Pred. No. 4.3e-69;
  Matches  534;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10609 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 10668

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10669 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 10728

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10729 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 10788

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10789 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 10848

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10849 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 10908

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10909 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 10968

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10969 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 11028

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11029 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 11088

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11089 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 11142


RESULT 3
AB032367
LOCUS       AB032367              64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
```

```
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
            Sciences, Kitasato University, Microbial chemistry; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .64957
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            101. .12019
                     /gene="aveA1"
     CDS             101. .12019
                     /gene="aveA1"
                     /note="contains a loading module, Ave module 1 and module
                     2
                     multifunctional polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="type I polyketide synthase AVES 1"
                     /protein_id="BAA84474.1"
                     /db_xref="GI:5902891"
                     /translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
                     ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
                     AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
                     VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTR
                     MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
                     MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
                     RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
                     DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
                     LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
                     GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
                     MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
                     GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
                     GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
                     RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
                     GDPIEAQALLATYGQQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
                     LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
                     AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAG
                     LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
                     EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
                     VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
                     AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
                     AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
                     ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
                     PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
                     THHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
                     EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
                     WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
                     LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
                     PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
                     LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGAL
                     ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
                     QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
                     TAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
                     DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSA
                     ARAREREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
                     FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
                     AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
                     YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
                     GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
                     GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
                     GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
                     NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
                     WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
                     TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
                     SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
                     ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
                     YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
                     VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
                     MTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
                     LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
                     VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
                     NLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
                     TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY
                     VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGASGG
                     VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
                     RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
                     EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
                     RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
                     LRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPE
                     DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
                     ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
                     VSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
                     AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
                     LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
                     RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
                     RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
                     TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
                     AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
                     SLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLGLS"
gene                 12071. .30790
                     /gene="aveA2"
CDS                  12071. .30790
                     /gene="aveA2"
                     /note="contains Ave module 3, module 4, module 5 and
                     module 6
                     multifunctional polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="type I polyketide synthase AVES 2"
                     /protein_id="BAA84475.1"
                     /db_xref="GI:5902892"
                     /translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
                     CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
                     DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGINP
                     EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
                     ACQALRSGECTMALAGGASVMATPFVFTEFSRQRGLAADGRCKAFSAAADGTGWSEGV
                     GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
                     PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
                     GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
                     GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
                     KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
                     ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
                     NDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALHRL
                     LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
                     TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
                     FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
                     TQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
                     TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
                     PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
                     NRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLTAVPADEPIAIV
                     GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
                     FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
                     LTYHDYAARFPTAPAGFEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
                     HLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSAAADGTGWGE
                     GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAD
                     LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
                     VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
                     SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAADDVPGEAAGDDAGTGGEATGPA
                     AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
                     IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH
                     GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
                     PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
                     QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
                     QGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
                     TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
                     PQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
                     SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
                     QHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
                     PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
                     LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
```

IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG
TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
GTWQGNGLATGQVSEHLRRRGMFAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA
AVLGRDDSEAIPPGRLFRDLGFDSLAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ
FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI
GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAAL
LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
TLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
KATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD
IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL

Query Match 100.0%; Score 534; DB 1; Length 64957;
Best Local Similarity 100.0%; Pred. No. 3.7e-69;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10709 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 10768

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10769 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 10828

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10829 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 10888

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10889 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 10948

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10949 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 11008

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11009 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 11068

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11069 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 11128

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      11129 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 11188

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11189 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 11242
```

RESULT 4
E38021
LOCUS E38021 31422 bp DNA linear PAT 31-JAN-2002
DEFINITION Avermectin aglycon synthase gene.
ACCESSION E38021
VERSION E38021.1 GI:18626910
KEYWORDS JP 2000245457-A/2.
SOURCE Streptomyces avermitilis.
ORGANISM Streptomyces avermitilis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 31422)
AUTHORS Omura,S. and Ikeda,H.
TITLE Avermectin aglycon synthase gene
JOURNAL Patent: JP 2000245457-A 2 12-SEP-2000;
THE KITASATO INSTITUTE
COMMENT OS Streptomyces avermitilis
PN JP 2000245457-A/2
PD 12-SEP-2000
PF 24-FEB-1999 JP 1999046961
PR
PI SATOSHI OMURA,HARUO IKEDA
PC C12N15/00,A61K31/70,C12N1/15,C12N9/88,C12P19/62, PC
C12Q1/68//C07D493/22,
PC (C12N1/15,C12R1:465),(C12N9/88,C12R1:465),C12N15/00 CC
FH Key Location/Qualifiers
FT CDS (1). .(14643)
FT CDS (14824). .(31419).
FEATURES Location/Qualifiers
source 1. .31422
/organism="Streptomyces avermitilis"
/db_xref="taxon:33903"
BASE COUNT 4136 a 10237 c 11677 g 5372 t
ORIGIN

Query Match 99.7%; Score 532.4; DB 6; Length 31422;
Best Local Similarity 99.8%; Pred. No. 7.3e-69;
Matches 533; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19285 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 19344

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19345 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 19404

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      19405 GGTGCGGAGGGTTTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 19464

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19465 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 19524

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19525 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 19584

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19585 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 19644

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19645 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 19704
```

```
Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19705 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 19764

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19765 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 19818


RESULT 5
AB032367/c
LOCUS       AB032367               64957 bp    DNA     linear   BCT 18-SEP-1999
DEFINITION  Streptomyces avermitilis polyketide synthase gene cluster (aveA1,
            aveA2, aveA3, aveA4) and aveC, aveE genes, complete cds.
ACCESSION   AB032367
VERSION     AB032367.1  GI:5902890
KEYWORDS    AveC; cytochrome P450 hydroxylase; type I polyketide synthase AVES
            4; type I polyketide synthase AVES 3; type I polyketide synthase
            AVES 2; type I polyketide synthase AVES 1.
SOURCE      Streptomyces avermitilis DNA.
  ORGANISM  Streptomyces avermitilis
            Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
            Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Organization of the biosynthetic gene cluster for the polyketide
            anthelmintic macrolide avermectin in Streptomyces avermitilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9509-9514 (1999)
  MEDLINE   99380548
REFERENCE   2  (bases 1 to 64957)
  AUTHORS   Ikeda,H., Nonomiya,T., Usami,M., Ohta,T. and Omura,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Haruo Ikeda, School of Pharmaceutical
            Sciences, Kitasato University, Microbial chemistry; 5-9-1
            Shirokane, Minato-ku, Tokyo 108-8641, Japan
            (E-mail:ikeda@mc.pharm.kitasato-u.ac.jp, Tel:+81-3-5791-6242,
            Fax:+81-3-3444-6197)
FEATURES             Location/Qualifiers
     source          1. .64957
                     /organism="Streptomyces avermitilis"
                     /db_xref="taxon:33903"
     gene            101. .12019
                     /gene="aveA1"
     CDS             101. .12019
                     /gene="aveA1"
                     /note="contains a loading module, Ave module 1 and module
                     2
                     multifunctional polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="type I polyketide synthase AVES 1"
                     /protein_id="BAA84474.1"
                     /db_xref="GI:5902891"
                     /translation="MQRMDGGEEPRPAAGEVLGVADEADGGVVFVFPGQGPQWPGMGR
                     ELLDASDVFRESVRACEAAFAPYVDWSVEQVLRDSPDAPGLDRVDVVQPTLFAVMISL
                     AALWRSQGVEPCAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTGALVS
                     VAATPDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSGAREAVADLVADLTAAQVRTR
                     MIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAYWYRN
                     MSSTVRFEPAARLLLQQGPKTFVEMSPHPVLTMGLQELAPDLGDTTGTADTVIMGTLR
                     RGQGTLDHFLTSLAQLRGHGETSATTVLSARLTALSPTQQQSLLLDLVRAHTMAVLND
                     DGNERTASDAGPSASFAHLGFDSVMGVELRNRLSKATGLRLPVTLIFDHTTPAAVAAR
                     LRTAALGHLDEDTAPVPDSPSGHGGTAAADDPIAIIGMACRFPGGVRSPKDLWELAAS
                     GGDAIGPFPTDRGWPTEQRHAQDPTQPGTFYPQGGGFLHDAAHFDAGFFGISPREALA
                     MDPQQRLLLETSWEAFERAGIDPLSVRGSRTGVFAGALSFDYGPRMDTASSEGAADVE
                     GHILTGTTGSVLSGRIAYSFGLEGPAITVDTGCSASLVTLHLACQSLRSGECTLALAG
                     GVSVMSTLGMFIEFSRQRGLSVDGRCKAYSAAADGTGWGEGVGMLLVERLSDAVRLGH
                     RVLAVVRGSAVNQDGASNGLTAPNGPAQERVIRQALANAGLSVADVDVVEGHGTGTTL
                     GDPIEAQALLATYGQRAGDRPLWLGSLKSNIGHTMAAAGVGGVIKMVMALREGVLPRT
                     LHVDKPSPQVDWSAGAVRLLTEAVPWPGDAAGRLRRAGVSSFGIGGTNAHVILEEAPA
                     AGGCVAGGGVLEGAPGLAISVAESVAAPVAVSAPVAESVPVPVPVPVPVPVSARSEAG
                     LRAQAEALRQYVAVRPDVSLADVGAGLACGRAVLEHRAVVLAADREELVQGLGALAAG
                     EPDRRVTTGHAPGGDRGGVVFVFPGQGGQWAGMGVRLLASSPVFARRMQACEEALAPW
                     VDWSVVDILRRDAGDAVWERADVVQPVLFSVMVSLAALWRSYGIEPDAVLGHSQGEIA
                     AAHVCGALSLKDAAKTVALRSRALAAVRGRGGMASVPLPAQEVEQLIGERWAGRLWVA
                     AVNGPRSTAVSGDAEAVDEVLAYCAGTGVRARRIPVDYASHCPHVQPLREELLELLGD
                     ISPQPSGVPFFSTVEGTWLDTTTLDAAYWYRNLHQPVRFSDAVQALADDGHRVFVEVS
                     PHPTLVPAIEDTTEDTAEDVTAIGSLRRGDNDTRRFLTALAHTHTTGIGTPTTWHHHY
                     THHHTHPHPHTHLDLPTYPFQHQHYWLESSQPGAGSGSGAGAGSGAGSGRAGTAGGTA
                     EVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRHQHDQARINTWTYQET
                     WKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGITPIPLTLNHTHTNPQH
                     LHHTLHHTRQQAQNHTTGAITGLLSLLALDETPHPHHPHTPTGTLLNLTLTQTHTQTH
                     PPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGIIDLPTTPTPHT
                     LQHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHGTTLITGGTGAL
                     ATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTITTCDTSNPDQLQ
                     QLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAHLLHQLTQHTPL
                     TAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAWGTWQGNGLADS
                     DKARAYLDRRGFRPMSPELATAAVTQAIADTERPYVVIADIDWSKIEHTSQTSDLVSA
                     ARELREPAVQRPTPPAELHKTLAHQTSADQRAALLELVRDHVAAVLRHADPKAIAPDQS
                     FRALGFDSLTAVEFRNLLIKATGLRLPVSLVFDHPTPAKLAVHLQNQLRGTAAESAPS
                     AAAVTAEASVTEPIAIVGMACRFPGGVTSADDFWDLISSEQDAIGGFPTDRGWDLDTL
                     YDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHA
                     GINPHTLHGTPTGVFTGTNGQDYALRVHNAGQSTDGFALTGTAGSVISGRISYTFGFE
                     GPAVSVDTACSSSLVALHLACQALRAGECSMALAGGVTVMSSPGAFVEFSRQRGLAAD
                     GHCKAFSAAADGTGWGEGVGMLLVERLSDAHRNGHRVLAVVRGSAVNQDGASNGLTAP
                     NGPSQQRVIRQALANAGLSAGDVDAVEAHGTGTTLGDPIEAQALLATYGQDRAGEGPL
                     WLGSVKSNVGHTQAAAGVAGVIKMVMALRHGLLPRTLHVDEPSPHVDWSAGAVQLLTE
                     TVPWPGGEGRLRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPAGEGDAGSDDEAAAG
                     SPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIA
                     ADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGL
                     YHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNEDAAALLQQTRYAQPALFAFQ
                     VALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGT
                     MTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKT
                     LPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNT
                     VDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNPPTTTLTLTHPHHHPQTHLLT
                     NLAKTTTTWHPHHYTHHDNQPHTHTHLDLPTYPFQHHHYWLESTQPGAGNVSAAGLDP
                     TEHPLLGATLELATDGGALLAGRLSLRSHPWLADHAVGGTVLLSGATFLELALHAGTY
                     VGCDRVDELTLHAPLVVPVDGGVSVQVGVAAADGEGRRLVSVYARGGSACGGGASGG
                     VWTCHASGVLVEAAAGGVVVDGLAGVWPPRGAVAVDVDGVRDRLAGAGCVLGPVFSGL
                     RAVWRDGGDLLAEVCLPEEAWGDAAGFGLHPALLDGVVQPLSVLLPGGTGFGEGAGFG
                     EGVRVPAVWGGVSLHRAGVTGVRVRVSAVGRGGGREAVSVVVGDEAGVPVASVDRLEL
                     RPVDMGQLRAVSVSAGRRGSLYAVQWAEVGPVPVCGQAWAWHEDVGESGGGPVPGVVV
                     LRCPDAGAGGGGGGGGGGVGEVVGGVLGVVQGWLGLERFAGSRLVVVTRGAVVAGPE
                     DGPVDVVGASVWGLVRSAQAEHPDRFVLLDLDTDTGTDLDTGAGAGWGVDGGRVAAVV
                     ACGEPQLAVRGERLLAARLKRLESSGDVPAQRSGDTRARRSDVPAQRSGGVPARRSVD
                     VSGREVLPWLSGGSVLVTGGTGVLGAAVARHLAGVCGVRDLLLVSRRGPDAPGAEGLR
                     AELAALGAEVRIVACDVGERREVVRLLEGVPAGCPLTGVVHAAGVLDDATIASLTPER
                     LGTVFAAKVDAALLLDELTRGMELSAFVLFSSAAGILGSAGQGNYAAANAALDALAYR
                     RRAAGLPGVSLAWGLWEEASGMTGHLAGTDHRRIIRSGLHPMSTPDALALFDAALALD
                     RPVLLPADLRPAPPLPPLLQDLLPATRRRTTRTTTTGGADNGAQLHARLAGQTHEQQH
                     TTLLALVRSHIATVLGHTTPDTIPPDRAFRDLGFDSLTAVELRNRLSRTTGLRLPTTL
                     AFDHPNPTTLTHHLHTQLQPQPDNAVAPVLAELDKLESALSALDKTDSASERVTLRLK
                     SLMLRWNAPQHPTAESADDDEKFTSATEAEIFKFIDNDLGLS"
gene                 12071. .30790
                     /gene="aveA2"
CDS                  12071. .30790
                     /gene="aveA2"
                     /note="contains Ave module 3, module 4, module 5 and
                     module 6
                     multifunctional polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="type I polyketide synthase AVES 2"
                     /protein_id="BAA84475.1"
                     /db_xref="GI:5902892"
                     /translation="MQLANEAKLLEYLKRVTADLDRTRRRLYEVVEREQEPIAIVGMA
                     CRYPGGATSPTRLWHLVKSQTDAIGEFPTDRGWNLEQLYDPDPDRSGTSYTRSGGFLY
                     DAGDFDAAFFELSPREALAMDPQQRLLLETTWETFEQGGIDPRSMRGSRTGVFVGINP
                     EDYTTGYTHQPSNAVEGYLLTGSAASIASGRISYNFGLEGPAITIDTACSSSLVALHL
                     ACQALRSGECTMALAGGASVMATPFVFTEPSRQRGLAADGRCKAFSAAADGTGWSEGV
                     GMLLVERLSDARRNGHRVLAVVRGSAVNQDGASNGLTAPNGRSQVKVIRQALANAHLS
                     PADVDAVEAHGTGTTLGDPIEAQALVEAYGQDRPNGRPLWLGTLKSNIGHSMAAAGVG
                     GVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVSSF
                     GVSGTNAHVILEEAPAHNIPSDTPADDVPGESAADEDAGSGDEAAAGSPGVWPWLVSA
                     KSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATLIAADRDTFLQALQ
                     ALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAHGLYHTHPVFAAAL
```

NDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTRYAQPALFAFQVALHRL
LTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLMQTMPPGTMTTLHT
TPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQQGIKTKTLPTNHA
FHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYWTQQARNTVDYATT
TQTLHQHGVTTYIELGPDNTLTTLTHDNLPNTPTTTLTLTHPHHHPQTHLLTNLAKTT
TTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLQPPGKPSDPSPSEGREQATTPST
PLRDVLVGKSPQERDEELLRLVRTHAAAVLGHATPEVIVPNKAFKELGFDSLAAIQLR
NRLLADVDLPLPATLIFDYPTPMALCQFLRAAIVGADTGTTTRLPLTAVPADEPIAIV
GMACRYPGDVRTVDDLWQVVSGGHDAIGGFPTNRGWDLDTLYNPDPDHHGTSYTRSGG
FLYDAGNFDPDFFGISPREALAMDPQQRLLLETAWESIEHACINPDSLRGTPTGVFAG
LTYHDYAARFPTAPAGFEGYLGHGSAGSIASGRVAYALGLEGPALTVDTACSSSLVAL
HLACQALRSGECSMALAGGVTVMSTPAGFVEFSRQRGLAVDGRCKAFSAAADGTGWGE
GVGMLLVERLSDARRLGHRILAVVRGSAVNQDGASNGLTAPNGPSQERVIRLALANAD
LTPADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPGNEPLWLGSMKSNIGHAQAAAG
VGGVIKMVMALRNGLLPRTLHVDEPSPHVDWSAGAVQLLTETVPWPGGEGRLRRAGVS
SFGVSGTNAHVILEEAPAHNIPSDTPADDAPGEAAADDVPGEAAGDDAGTGGEATGPA
AGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAHARAVFDHRATL
IAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICSGQGTQRPGMAH
GLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQDPNTQDTTTLEEAAALLQQTPYAQ
PALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTLITQRATLM
QTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQHITTLCQQ
QGIKTKTLPTKNAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPDQLLTPHYW
TQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTLTLTHPHHH
PQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHQHYWLESTQPGAGSG
SGSGSGRAGTAGGTAEVESRFWDAVARQDLETVATTLAVPPSAGLDTVVPALSAWHRH
QHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHGIT
PIPLTLNHTHTNPQHLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTPTGTLLN
LTLTQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHPTHTAGI
IDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPPTPTPHG
TTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKGIHLTIT
TCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGILDDATLTNLTPTQLNNVLRAKAHSAH
LLHQLTQHTPLNAFVLYSSAAATFGAPGQANYAAANAYLDALAHHRHTHHLPATSIAW
GTWQGNGLATGQVSEHLRRRGMPAMPPELAVTAVDGAIASGRPSLLVADIDWKKLGPV
LSSKSSVLLEDLPQAQGTEEARSTVEQTESTNLRQLLMGRSRSEQEEELLSLVRIHSA
AVLGRDDSEAIPPGRLFRDLGFDSLAAAVELRNHLAAQTELALPTTLVFDYPSPTKLAQ
FLLSEIAEFQPDNSTPLPRPRAELDEPIAIVGMACRFPGGVTSADDFWDLISSEQDAI
GGFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFLYDAGHFDAEFFGISPREALAMDPQQ
RLLLETAWETIEHAGINPHTLHGTPTGVFTGTNGQDHAAHIRQAPSGTEGFVLTGAAT
SIASGRISYILGLEGPAVTLDTACSSSLVALHLACQSLRSGECTMALAGGATVMTTPI
TFTEFARQRGLAPDGRCKAFSAAADGTGWGEGVGMLLVERLSDARRNGHRVLAVVRGS
AVNQDGASNGLTAPNGPSQQRVIRQALANADLTPADVDAVEAHGTGTTLGDPIEAQAI
LATYGQDRPGNGPLWLGSVKSNVGHTQAAAGVAGVIKMVMALRHRTLPPTLHADEPSP
HVDWSAGAVQLLTETVPWPGGEGRPRRAGVSSFGVSGTNAHVILEEAPADDVPGGPPA
DEDAGSGEEAAAGSPGVWPWLVSAKSQPALRAQAQALHAHLTDHPGLDLADVGYTLAH
ARAVFDHRATLIAADRDTFLQALQALAAGEPHPAVIHSSAPGGTGTGEAAGKTAFICS
GQGTQRPGMAHGLYHTHPVFAAALNDICTHLDPHLDHPLLPLLTQNDNDNDNEDAAAL
LQQTPYAQPALFAFQVALHRLLTDGYHITPHYYAGHSLGEITAAHLAGILTLTDATTL
ITQRATLMQTMPPGTMTTLHTTPHHITHHLTAHENDLAIAAINTPTSLVISGTPHTVQ
HITTLCQQQGIKTKTLPTNHAFHSPHTNPILNQLHQHTQTLTYHPPHTPLITANTPPD
QLLTPHYWTQQARNTVDYATTTQTLHQHGVTTYIELGPDNTLTTLTHHNLPNTPTTTL
TLTHPHHHPQTHLLTNLAKTTTTWHPHHYTHHHNQPHTHTHLDLPTYPFQHHHYWLEL
PSAQTSPGQRRSRRSAPDTAESEFWDAVNEEDLQSLAETLDIDASALDTVVPALSAWH
RHQHDQARINTWTYQETWKPLTLPTTHQPHQTWLIAIPETQTHHPHITNILTNLHHHG
ITPIPLTVNHTHTNPQHLHHTLHHTRQQAQNHTTGPITGLLSLLALDETPHPHHPHTP
TGTLLNLTLPQTHTQTHPPTPLWYATTNATTTHPNDPLTHPTQAQTWGLARTTLLEHP
THTAGIIDLPTTPTPHTLHHLTQTLTQPHHQTQLAIRTTGTHTRRLTPTTLTPTHQPP
TPTPHGTTLITGGTGALATHLTHHLTTHQPTQHLLLTSRTGPHTPHAQHLTTQLQQKG
IHLTITTCDTSNPDQLQQLLNTIPPQHPLTTVIHTAGVNLFAPVSETDAESFSSVTAA
KATGAAILHELLLDHETLEHFILFSSGAGAWGSGNQCAYSAANAYLDALATHRQTHGL
PGASIAWGPWAGKGMSAGDAAHGYLEKRGILPMEPRMALAAFHRARAQRPNSNLIIAD
IDWERFVPAFTARRHSPLIEDIPEVRQAAQELEAAASTAKTTTAQPIATSLRERLARL
TSSKQNQVLLGLIRTGICTVLGLRNPEGIEDQRAFRDLGFDSLTSAQFSKELAKETGL
PLPPSLVFDYPTPQECAAHLRTQLVDLDDEEDAALSNALPQVAHRRTVEDEPIAIIGM
ACRFPGGVRSADDLWELLASGKDAIGVFPTDRGWDLDTLYDPDPDHPGTCYTRNGGFL
YGAGHFDAEFFGISPREALAMDPQQRLLLETAWETIEHAGINPHTLHGTPTGVFAGIN
AQDHAAHIRQSRDVETIEGYALTGSSGSVASGRVAYTLGLEGPAVSVDTACSSSLVAL

Query Match 99.7%; Score 532.4; DB 1; Length 64957;
Best Local Similarity 99.8%; Pred. No. 6.3e-69;
Matches 533; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45573 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 45514

Qy         61 GCTGGTGTGTGTGGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45513 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 45454

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      45453 GGTGCGGAGGGTTTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 45394

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45393 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 45334

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45333 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 45274

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45273 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 45214

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45213 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 45154

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45153 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 45094

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      45093 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 45040
```

RESULT 6
SGR300302/c
LOCUS SGR300302 39314 bp DNA linear BCT 20-MAR-2002
DEFINITION Streptomyces griseus partial ORF1, canA gene, canC gene, canF gene, canT gene, canRA gene and canRB gene.
ACCESSION AJ300302
VERSION AJ300302.1 GI:12231153
KEYWORDS ABC-transporter; aminotransferase; can RB gene; canA gene; canC gene; canF gene; canRA gene; canT gene; Cho-like protein; cytochrome P-450; ferredoxine; PABA synthase; thioesterase.
SOURCE Streptomyces griseus.
ORGANISM Streptomyces griseus
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Campelo,A.B.
JOURNAL Thesis (2000) Department of Microbiologia, Universidad de Leon, Leon, Spain
REFERENCE 2
AUTHORS Campelo,A.B. and Gil,J.A.
TITLE The candicidin gene cluster from Streptomyces griseus IMRU 3570
JOURNAL Microbiology 148 (Pt 1), 51-59 (2002)
MEDLINE 21642576
PUBMED 11782498
REFERENCE 3 (bases 1 to 39314)
AUTHORS Gil,J.A.
TITLE Direct Submission
JOURNAL Submitted (14-NOV-2000) Gil J.A., Microbiologia, Universidad de Leon, Campus de Vegazana s/n, 24071, SPAIN
FEATURES Location/Qualifiers
source 1. .39314
/organism="Streptomyces griseus"
/strain="IMRU 3570"
/db_xref="taxon:1911"
CDS complement(<1. .1785)
/note="ORF1"
/codon_start=1
/transl_table=11

```
                     /product="Cho-like protein"
                     /protein_id="CAC22112.1"
                     /db_xref="GI:12231154"
                     /db_xref="SPTREMBL:Q9EWC6"
                     /translation="MLLERTTEIALVDAALGAATEGRSSLVLLTGPLGIGRSALLQRL
                     SCLADDRDDVRVLRAHAAPMEQDFAFGVVRQLPETLLGDSPEDARDRWMDAHASFARH
                     VLADDAAPPGADQAIAATEAVLHGLLSLLANVSADSRLLILVDDLQWSDVPSLRWLTY
                     LAKRLHGLRAVVVCALRDGDPRSHHTLVREIREAATQTLRPASLSLTATHELVREHFG
                     EAGDDEFVQACHEASVGNPLFLLSILVGTGFLGRRPLAEHAETARRLRPSQLRERLAS
                     ILRTQPAPVRDLAAAIAILGEQSDAPTLARLAGLDSIGYAGALRALGALGALAAPDEP
                     RFIHRSVRDAAESTLTMVQRERMHDEAAALLYAAGSPAEQVAAQLMAVVTRRQPWAVD
                     VLRAAADTALRRGAPDTAAGYLCRALLDSPAAGVGRGRLLVELGTAERGFDPLACERH
                     IAQAMTLLPEPRDRAVAALRISPTALGPAPLTAVDLLRQAAEDLGPADGLTGTDRDLA
                     LRLEARLRHCGHEDPHELAASAHRLKELGPEPPVDTGGERELTIALLGAATLGCHLPA
                     SEVVRLAGRVLDREPATSARIHSTLPLAI"
gene                 9525. .10583
                     /gene="canA"
CDS                  9525. .10583
                     /gene="canA"
                     /codon_start=1
                     /transl_table=11
                     /product="aminotransferase"
                     /protein_id="CAC22113.1"
                     /db_xref="GI:12231155"
                     /db_xref="SPTREMBL:Q9EWC5"
                     /translation="MAFTHPVSRPALDGRELEYVSDAVSGGWISSQGPYVRRFEEAFA
                     EWNGVAHGVACSSGTAALTLALRALNIGPGDEVIVPEFTMVASAWAVTYTGATPVFVD
                     CGDDLNIDVTRIEEKITPRTRAVMPVHVYGRRCDMDAVMDLALQYNLRVVEDSAEAHG
                     VRPVGDIACFSLFANKIITAGEGGVCLTDDPRLAEQLAHLRAMAFTRDHSFLHKKLAY
                     NYRMTAMQGAVALAQTERLDEILATRREIEARYDAGLKDLPGITLMPARDVLWMYDLR
                     AERREELRAHLDERGIETRLFFKPMSRQPGYLDPVWPTLNAHRFSEDGLYLPTHTGLT
                     AADQEYITGAVRDFYRAG"
gene                 10678. .11859
                     /gene="canC"
CDS                  10678. .11859
                     /gene="canC"
                     /codon_start=1
                     /transl_table=11
                     /product="cytochrome P-450"
                     /protein_id="CAC22114.1"
                     /db_xref="GI:12231156"
                     /db_xref="SPTREMBL:Q9EWC4"
                     /translation="MTTSPGPTVVDFPRRTPREPLPLSQYAEHRKQNGLVQTHLPNGR
                     PIWLVTRHEDVRAVLTHPRISANPDNEGFPNVGETMGVPKQEQIPGWFVGLDSPEHDR
                     FRKVLIPEFTVRRVRELRPAIERTVDERIDAMLAGGNTADLVNDFALPVPSLVISALL
                     GVPSADRDFFESRTRTLVAIRTSTDEERAEATRQLLRYINRLIVIKKKWRGEDLISRL
                     LSTGKLSDEELSGVLLLLLIAGHETTANNIGLGVVTLLSHREWIGDDRLVEELLRLHS
                     VADMVALRVAVDDVEIAGQTIRKGEGIVPLLASANHDTEAFGCPHAFNPERTERRHVA
                     FGYGVHQCLGQNLVRVEMEIAYRKLFERIPELRLAVPEDQLAYKYDGILFGLHELPVR
                     W"
gene                 11906. .12100
                     /gene="canF"
CDS                  11906. .12100
                     /gene="canF"
                     /codon_start=1
                     /transl_table=11
                     /product="ferredoxine"
                     /protein_id="CAC22115.1"
                     /db_xref="GI:12231157"
                     /db_xref="SPTREMBL:Q9EWC3"
                     /translation="MRVTVDSEQCVGAGQCVLNAPEVFDQDDDGVVVLLRAEPDERDH
                     EAVRTAGDLCPSASVVLQED"
gene                 12223. .12993
                     /gene="canT"
CDS                  12223. .12993
                     /gene="canT"
                     /codon_start=1
                     /transl_table=11
                     /product="thioesterase"
                     /protein_id="CAC22116.1"
                     /db_xref="GI:12231158"
                     /db_xref="SPTREMBL:Q9EWC2"
                     /translation="MTTADDTAARWLRRYHPAEADAVRLVCFPHAGGSASFYHPVSAR
                     FAPGAEVVSLQYPGRQDRRKEPCVPDLGTLADLITEQLLPLDERPTVFFGHSMGAALA
                     FETAWRLEQKGAGPRTVIASGRRGPSTTRAERVHTRDDDGIVAEMKRLNGTAAGVLGD
                     EEILRMALPALRGDYRAIETYTCPPDRRLACGLTVLTGEDDPLTTVEEAERWRDHTTG
                     PFRLRVFTGGHFFLTQHLDAVNTEIAQALHPDRAAPAA"
gene                 13190. .15361
                     /gene="pabAB"
CDS                  13190. .15361
                     /gene="pabAB"
                     /codon_start=1
                     /transl_table=11
                     /product="PABA synthase"
                     /protein_id="CAC22117.1"
                     /db_xref="GI:12231159"
                     /db_xref="SPTREMBL:Q9EWC1"
                     /translation="MRTLLVDNYDSFTYNLFHYLSRANGREPEVIRNDDPAWRPGLLD
                     AFDNVVLSPGPGTPHRPADFGLCARIAEEGRLPVLGVCLGHQGMALAHGARVGRAPEP
                     RHGRTSAVRHDGTGLFEGLPQPLEVVRYHSLAVTELPPELEATAWSEDGVLMALRHRT
                     LPLWGVQFHPESIGTQDGHRLLANFRDLTERHGRTRHGGRAGHGTLPPPAPARETKAT
                     TGTPRRLRVIAKSLPTRWDAEVAFDSLFRTGDHPFWLDSSRPGGELGQLSMMGDASGP
                     LARTAKADVHAGTVTVRADGASSTVESAFLTWLENDLAGLRTEVPELPFAFALGWVGC
                     LGYELKAECDGDAAHRSPDPDAVLVFADRALVLDHRTRTTYLLALVEDDAEAEARAWL
                     AAASATLDAVAGREPEPCPEAPVCTTGPVELRHDRDGYLKLIDVCQQEIAAGETYEVC
                     LTNMAEADTDLTPWAAYRALRRVSPAPFAAFLDFGPMAVLSSSPERFLRIDRHGRMES
                     KPIKGTRPRGATPQEDAALVRALATCEKDRAENLMIVDLVRHDLGRCAEVGSVVADPV
                     FQVETYATVHQLVSTVTARLREDSSPVAAVRAAFPGGSMTGAPKIRTMQIIDRLEGGP
                     RGVYSGAIGYFSLTGAVDLSIVIRTVVLSGGRLRYGVGGAVIALSDPADEFEETAVKA
                     APLLRLLDTAFPGREAPGKDLDGEPDDGTDAGAPKDLVLPG"
gene                 20829. .21836
                     /gene="canRA"
CDS                  20829. .21836
                     /gene="canRA"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC-transporter"
                     /protein_id="CAC22118.1"
                     /db_xref="GI:12231160"
                     /db_xref="SPTREMBL:Q9EWC0"
                     /translation="MNGLHLPPAPLPAPGVSAAGSGDLMIEAEGVSKAYGTVHALDSV
                     SLAVPRGTVLGLLGHNGAGKTTLVDILTTALPPTSGRARVAGYDVADKPVEIRRRIGL
                     TGQFASVDAQLSGYDNLVLIARLLGAGRRAARVRARELLELFRLTEVADRRASSYSGG
                     LRRRLDLAVSLVGAPEVLFLDEPTTGLDPSSRINLWEIVEGLVEQGTTVLLTTQYLEE
                     ADRLADRIAVLSAGRVVAAGTAPELKATVGRRTVTLTLETGEDVSAARTALRGAGFAP
                     VDGEARTVVVPIDATREIADIIRSLDQAGVEASELNFGEPTLDDVYLTLAEQAAGRTR
                     A"
gene                 21970. .22776
                     /gene="canRB"
CDS                  21970. .22776
                     /gene="canRB"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC-transporter"
                     /protein_id="CAC22119.1"
                     /db_xref="GI:12231161"
                     /db_xref="SPTREMBL:Q9EWB9"
                     /translation="MPRPAAVPAFGGSSLFTQIRVLTGRSLRAMVTDPGIVLFGLIQP
                     VVILFVLTQVFSKMGMPPHFPDGVSYLDYVLPAVLVDNAAQSAMQSGVGLVEDQKNGI
                     VARLRSLPVHPGALLAARSLVGLVRSAVQVAVIMALAMTVLGYSPNGGAAELALSAGL
                     TLFISFSLGWAFIAAGAWLRRAEPLQNLALIVIFPLMFASSAYVPVADLPSWLGAVAS
                     VNPLTYAIDATRALALDVPGLDHAIPAVVICAVIAVVGMIAAVRGFRRPL"
BASE COUNT     5144 a  14526 c  14888 g   4756 t
ORIGIN

  Query Match           48.5%;  Score 258.8;  DB 1;  Length 39314;
  Best Local Similarity   67.8%;  Pred. No. 4.5e-29;
  Matches  362;  Conservative    0;  Mismatches  172;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  |||||  |||||||||||||  |||||||  ||||||||||  | ||||||||  |||||
Db      35608 GGTGCGGTGCTGGTGACGGGTGGTACGGGTGGTCTGGGTGCGGTGATGGCGCGGTATCTG 35549

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              | | || | ||||||||||||||||||  || |||||   |||||  ||   |||  ||||||
Db      35548 GTGGCTGAGCGTGGGGTGCGGGATCTGGTGCTGGTGTCGCGGCGGGGCGGGGACGCTCCG 35489
```

```
Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              |||||| ||  |||    | ||||||   | |    || |||| ||||  | | || |||
Db      35488 TGTGCGGCGGAGCTGGCCGGAGAGCTGCGTGAGGCCGGTGCGGCGGTGGAGGTCGTCGCG 35429

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              || ||  ||   || ||      ||| ||| | ||| | |||    ||   |||||| ||
Db      35428 TGCGACCTGTCCGACCGCGAATCGGTCGTCGGCCTGGTCGAGTCCCTTGTCGCCGGGCGT 35369

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
                ||||   | |||||||| ||||| || ||    | || |    | | | | | ||
Db      35368 GGGCTGCGCGCCGTCGTGCACGCGGCCGGGGTCGGCGGCGGCGGCCTGGTGGGCACGCTG 35309

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              || |||| ||| | |   ||||| | |  || ||||  ||||| ||    | | || |
Db      35308 TCGTCCGACCGGTTCGATGCGGTGCTGGGTGCGAAGGCTGATGCGGCGTGGTGGTTGCAC 35249

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||  ||||  |||| |||||||| | |||||||||||| | || |||||  |  ||  |
Db      35248 GAGGCGACGGCGGGTGTGGAGCTGGCTGCGTTCGTGCTGGTGTCGTCGGCGGGTGGTCTG 35189

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              ||  | |||| |||||||||||||||||| ||||||||||     |||||||||| |||||
Db      35188 GTGCTGACGGCGGGGCAGGGCAACTACGCGGCGGCCAATGTGTTCCTGGACGCGTTGGCG 35129

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                |||||  |||||  | ||| |||  ||      ||  ||||||  ||| | |||
Db      35128 GCGCGGCGTCGGGCCGGGGGCCTGGTGGCCACATCCATGGCGTTCGGGTTCTGG 35075


RESULT 7
AX211733
LOCUS       AX211733               2700 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 29 from Patent WO0159126.
ACCESSION   AX211733
VERSION     AX211733.1  GI:15523945
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 2700)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 29 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .2700
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT      324 a   1087 c    876 g    413 t
ORIGIN

  Query Match             48.2%;  Score 257.2;  DB 6;  Length 2700;
  Best Local Similarity   67.6%;  Pred. No. 1.3e-28;
  Matches  361;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  |||||||||||||||||| |||||||  |||||||  |  ||||||  |||||||||
Db       1170 GGAACGGTGTTGGTGACGGGTGGTACGGGTGGTCTGGGTGGGTTGGTGGCCCGGCATCTG 1229

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |   |   |||||||||||||| || || |||||    || ||   ||||  || |||| ||
Db       1230 GTGCGGTCGTGTGGGGTGCGGCATTTGTTGTTGACCAGTCGTTCTGGTGTGGGTGCTGCG 1289

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              |||||||  ||  ||    ||||||  |||  |||||||  |||   |||    | |||  |||
Db      1290 GGTGCGGCCGGGTTGGTCGCGGAGTTGGAGTCGTTGGGCGCGCGGGTTGTGGTTGCGGCG 1349

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||| ||   |  ||  ||  |    | ||  |  |||||  | |    || |
Db      1350 TGTGATGTGGGTGATGGCTCGGCTGTTGCGGAGTTGGTTGCCGGTGTGTCGGAGTCGTAT 1409

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||| || | |  || |||||||||||||||||| |||| || |    | | |  ||  | |
Db      1410 CCGTTGTCTGCGGTGGTGCATGCGGCTGGTGTGTTGGATGACGGTGTGGTGGGTTCGTTG 1469

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||||| |||||| |||    |||||||      | ||||||||||  ||        ||| ||
Db      1470 ACGCCGGAGCGGTTGGCTGCGGTGTTGCGTCCGAAGGTGGATGGTGCGTGGAACCTGCAT 1529

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||  |||||| ||| |||| |||    ||||| ||  | ||||| || |  ||||  |
Db      1530 GAGGCGACGCGTGGTCTGGATCTGGACGCGTTTGTTGTCTTCTCGTCTGTTGCGGGTGTG 1589

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | |||   || || |||| |||||| || |||| ||||   | ||||||| ||  |
Db      1590 TTCGGGGGTGCGGGTCAGGCCAACTATGCGGCGGGTAATGCGTTTTTGGACGCGTTGATG 1649

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                |  |||  ||| | ||  |||| || |||||| ||||||||||     |||
Db      1650 GTTCATCGGGTGGCTGGTGGGTTGCCTGGTGTGTCGTTGGCGTGGGGTGCTTGG 1703


RESULT 8
AF420594
LOCUS       AF420594               2706 bp    DNA     linear   BCT 25-OCT-2001
DEFINITION  Streptomyces noursei rearranged polyketide synthase (nysCd48) gene,
            partial cds.
ACCESSION   AF420594
VERSION     AF420594.1  GI:16417583
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 2706)
  AUTHORS   Brautaset,T., Bruheim,P., Sletta,H., Hagen,L., Ellingsen,T.E.,
            Strom,A.R., Valla,S. and Zotchev,S.B.
  TITLE     Hexaene derivative of the polyene antifungal antibiotic nystatin is
            produced as a result of an induced rearrangement within the nysC
            polyketide synthase gene in Streptomyces noursei ATCC 11455
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2706)
  AUTHORS   Brautaset,T. and Zotchev,S.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2001) Biotechnology, NTNU, S. Saelands v. 6/8,
            Trondheim 7491, Norway
FEATURES             Location/Qualifiers
     source          1. .2706
                     /organism="Streptomyces noursei"
                     /strain="ERD48"
                     /db_xref="taxon:1971"
     gene            <1. .>2706
                     /gene="nysCd48"
     CDS             <1. .>2706
                     /gene="nysCd48"
                     /note="NysCd48"
                     /codon_start=1
                     /transl_table=11
                     /product="rearranged polyketide synthase"
                     /protein_id="AAL18806.1"
                     /db_xref="GI:16417584"
                     /translation="GSADDTGRRTVTVHARPDDTADRTWTLHATGVLATTPPAAAAFD
                     TTVWPPADAEPLTTDDCYAHFTTHGFAYGPAFQGLRAAWRAGDVLYAEVALPESATDE
                     AAAFGLHPALLDAGLHAALLADDRDTGLPFSWEGVTLHASGATALRVRLAPNGPNGLS
                     VTAADPAGNPVATVTRLLARPLDAEQLTIHSAPTRDALFHLDWTPVPLPDTANSAPPA
```

```
                     LLGPDTAVLADALGDPAVARHATLDDLLAGDTTPPATVLVPLGAPLDGDTAQHAHALT
                     RSALTLVQQWLATDRLADSRLVFVTHGAVATDDAPPTDLAAAAVWGLIRSAQTENPGT
                     FTLLDLDTEPDSTTALSRALTLDEPQLLLRAGRARAARLTRTPAPTTTTHTPWSADGT
                     VLVTGGTGGLGGLVARHLVRSCGVRHLLLTSRSGVGAAGAAGLVAELESLGARVVVAA
                     CDVGDGSAVAELVAGVSESYPLSAVVHAAGVLDDGVVGSLTPERLAAVLRPKVDGAWN
                     LHEATRGLDLDAFVVFSSVAGVFGGAGQANYAAGNAFLDALMVHRVAGGLPGVSLAWG
                     AWDQGVGMTAGLTERDVRRAAESGMPLLTVDQGVALFDAALATGSAALVPVRLDLAAL
                     RTRGDIAPLLRGLVKAPIRRAAATTPGDTGLAEQLTRLQRAERRDTLLALVRDQAAMV
                     LGHTSGDGVDPSRAFRDLGFDSLTAVELRNRIGAATGLRLPATAVFDYPTADALAAHL
                     LTELLGPDAESDPDEPGDPTAGPTDDPIVIIGMSCRFPGDIGSPEDLWRLLGDGADVV
                     TDFPTNRGWDLDNLYDPDPAHAGTSYARTGGFLHDAADFDADFFGMSPREAMATDSQQ
                     RLLLESSWEAIERAGIDPLTLRDSRTGVFAGVMYSGYGTRLDGAEF"
     misc_feature    <1. .255
                     /gene="nysCd48"
                     /note="Region: dehydratase domain DH4; module 4"
     misc_feature    1171. .1914
                     /gene="nysCd48"
                     /note="result of recombination between KR4- and KR5-
                     encoding regions within nysC; results in deletion of one
                     complete module; Region: hybrid ketoreductase domain
                     KR4/5"
     misc_feature    2011. .2232
                     /gene="nysCd48"
                     /note="Region: acyl carrier protein domain ACP"
     misc_feature    2296. .>2706
                     /gene="nysCd48"
                     /note="Region: ketosynthase domain KS"
BASE COUNT      327 a   1088 c    877 g    414 t
ORIGIN

  Query Match             48.2%;  Score 257.2;  DB 1;  Length 2706;
  Best Local Similarity   67.6%;  Pred. No. 1.3e-28;
  Matches  361;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  |||||||||||||||||||| |||||||  |||||||  |  |||||| |||||||||
Db       1171 GGAACGGTGTTGGTGACGGGTGGTACGGGTGGTCTGGGTGGGTTGGTGGCCCGGCATCTG 1230

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |   |  |||||||||||||| || || |||||   || ||   ||||  || |||| ||
Db       1231 GTGCGGTCGTGTGGGGTGCGGCATTTGTTGTTGACCAGTCGTTCTGGTGTGGGTGCTGCG 1290

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              |||||||  ||  ||   ||||||| |||  ||||||| |||  |||   | |||  |||
Db       1291 GGTGCGGCCGGGTTGGTCGCGGAGTTGGAGTCGTTGGGCGCGCGGGTTGTGGTTGCGGCG 1350

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||| ||  |   ||  || |    | || | |  |||||  | |    || |
Db       1351 TGTGATGTGGGTGATGGCTCGGCTGTTGCGGAGTTGGTTGCCGGTGTGTCGGAGTCGTAT 1410

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |||  || | |  || ||||||||||||||||||| |||| || |    | | |  ||  |
Db       1411 CCGTTGTCTGCGGTGGTGCATGCGGCTGGTGTGTTGGATGACGGTGTGGTGGGTTCGTTG 1470

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |||||  ||||||  |||   |||||||     | ||||||||||  ||       ||| ||
Db       1471 ACGCCGGAGCGGTTGGCTGCGGTGTTGCGTCCGAAGGTGGATGGTGCGTGGAACCTGCAT 1530

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||  ||||||  ||| |||| |||   ||||| ||  | |||||  || |  |||||  |
Db       1531 GAGGCGACGCGTGGTCTGGATCTGGACGCGTTTGTTGTCTTCTCGTCTGTTGCGGGTGTG 1590

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | |||   || || |||| |||||| || ||||  |||||   | ||||||| ||  |
Db       1591 TTCGGGGGTGCGGGTCAGGCCAACTATGCGGCGGGTAATGCGTTTTTGGACGCGTTGATG 1650

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              |  |||  ||| |  ||  |||| || |||||| ||||||||||     |||
Db       1651 GTTCATCGGGTGGCTGGTGGGTTGCCTGGTGTGTCGTTGGCGTGGGGTGCTTGG 1704


RESULT 9
AX211705
```

```
LOCUS       AX211705               65140 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 1 from Patent WO0159126.
ACCESSION   AX211705
VERSION     AX211705.1  GI:15523937
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 65140)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 1 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .65140
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT     8270 a  25171 c  22273 g   9426 t
ORIGIN

  Query Match             48.2%;  Score 257.2;  DB 6;  Length 65140;
  Best Local Similarity   67.6%;  Pred. No. 6.9e-29;
  Matches  361;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  |||||||||||||||||||| |||||||  |||||||  |  |||||| |||||||||
Db      25919 GGAACGGTGTTGGTGACGGGTGGTACGGGTGGTCTGGGTGGGTTGGTGGCCCGGCATCTG 25978

Qy         61 GCTGGTGTGTGTGGGGTCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |   |  |||||||||||||| || || |||||   || ||   ||||  || |||| ||
Db      25979 GTGCGGTCGTGTGGGGTGCGGCATTTGTTGTTGACCAGTCGTTCTGGTGTGGGTGCTGCG 26038

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              |||||||  ||  ||   ||||||| |||  ||||||| |||  |||   | |||  |||
Db      26039 GGTGCGGCCGGGTTGGTCGCGGAGTTGGAGTCGTTGGGCGCGCGGGTTGTGGTTGCGGCG 26098

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||| ||  |   ||  || |    | || | |  |||||  | |    || |
Db      26099 TGTGATGTGGGTGATGGCTCGGCTGTTGCGGAGTTGGTTGCCGGTGTGTCGGAGTCGTAT 26158

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              |||  || | |  || ||||||||||||||||||| |||| || |    | | |  ||  |
Db      26159 CCGTTGTCTGCGGTGGTGCATGCGGCTGGTGTGTTGGATGACGGTGTGGTGGGTTCGTTG 26218

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |||||  ||||||  |||   |||||||     | ||||||||||  ||       ||| ||
Db      26219 ACGCCGGAGCGGTTGGCTGCGGTGTTGCGTCCGAAGGTGGATGGTGCGTGGAACCTGCAT 26278

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||  ||||||  ||| |||| |||   ||||| ||  | |||||  || |  |||||  |
Db      26279 GAGGCGACGCGTGGTCTGGATCTGGACGCGTTTGTTGTCTTCTCGTCTGTTGCGGGTGTG 26338

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | |||   || || |||| |||||| || ||||  |||||   | ||||||| ||  |
Db      26339 TTCGGGGGTGCGGGTCAGGCCAACTATGCGGCGGGTAATGCGTTTTTGGACGCGTTGATG 26398

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              |  |||  ||| |  ||  |||| || |||||| ||||||||||     |||
Db      26399 GTTCATCGGGTGGCTGGTGGGTTGCCTGGTGTGTCGTTGGCGTGGGGTGCTTGG 26452


RESULT 10
AF263912
```

```
LOCUS       AF263912              123580 bp    DNA     linear   BCT 24-MAY-2000
DEFINITION  Streptomyces noursei ATCC 11455 nystatin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF263912
VERSION     AF263912.1  GI:8050835
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., StrLm,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Biosynthesis of the polyene antifungal antibiotic nystatin in
            Streptomyces noursei ATCC 11455: analysis of the gene cluster and
            deduction of the biosynthetic pathway
  JOURNAL   Chem. Biol. 7 (6), 395-403 (2000)
  MEDLINE   20334850
   PUBMED   10873841
REFERENCE   2  (bases 1 to 123580)
  AUTHORS   Brautaset,T., Sekurova,O.N., Sletta,H., Ellingsen,T.E., Strom,A.R.,
            Valla,S. and Zotchev,S.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-MAY-2000) Unigen, NTNU, O. Kyrres gt. 3, Trondheim
            N-7489, Norway
FEATURES             Location/Qualifiers
     source          1. .123580
                     /organism="Streptomyces noursei"
                     /strain="ATCC 11455"
                     /db_xref="ATCC:11455"
                     /db_xref="taxon:1971"
     gene            complement(46. .783)
                     /gene="nysF"
     CDS             complement(46. .783)
                     /gene="nysF"
                     /function="presumably post-translationally modifies ACP
                     domains on PKS"
                     /note="putative 4'-phosphopantheteine transferase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysF"
                     /protein_id="AAF71762.1"
                     /db_xref="GI:8050836"
                     /translation="MIELILPATVATEAAYDDRPRPGDRLLSSEREVIARAVESRQRE
                     FTTVRHLARRALRRLGHPDRAILPNRRGAPQWPPGIVGSMTHCAGYRAAAVSPAELSA
                     AVSIDAEPNGPLPAGVLNAIALPSERPHLVALAAHRPDVHWDRLLFSAKESVFKAWYP
                     LTQRELDFSEAEIVIDPTQGAFTARLLVPGPLLGGRRVTVFPGRWHSTPALLTTAVHL
                     PAPTPRRDREHRTHLTVNSPLPRPTFG"
     gene            complement(867. .2684)
                     /gene="nysG"
     CDS             complement(867. .2684)
                     /gene="nysG"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysG"
                     /protein_id="AAF71763.1"
                     /db_xref="GI:8050837"
                     /translation="MASPDDLEEERTAPRPARRLVGLLRPHRRSVALAVSMGVGGIVL
                     NAFGPLLLGRVTDLIADGVLGGVPGPAPGIDFAAIGRLLLVLLALYVVASLFMLAQGR
                     LVASAVWRTIHELRRDAREKLTRLPLRHFDRQPAGELLSRTTNDIDNLQQTLQQTLAE
                     LITSIFSLLTMLVLMLVISPSLAVVMLLSVPVSALIAARISKRAQPHYAAQWSANGTL
                     NAHVEEVCTGHALIKGFDRRAAAEERFDACNDAVYRAAAKAQFASGAMEPVMMFVANL
                     GYVLVAVIGAWKVINGTLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGIASAQRVF
                     TLLDAPEQAPDPLRPGTPARAEGRVEFTDVSFRYSPDTPLIENLSLTVEPGSTVAIVG
                     PTGAGKTTLGNLLMRFYEPDSGRILLDGTDTATMTRDDLRSRFGLVLQDTWLFGGTIA
                     ENIAYGAPGACRADIEEAARATCADRFIRTLPQGYDTVLDDESGTVSAGEKQLLTVAR
                     AFLARPAVLVLDEATSSVDTRTEVLIQRAMNSLRAGRTSFVIAHRLSTIRDADLIVVM
                     DAGRIVEQGTHDQLLCAQGLYARLHAARTHTPTAGAAAG"
     gene            complement(2662. .4416)
                     /gene="nysH"
     CDS             complement(2662. .4416)
                     /gene="nysH"
                     /function="presumably involved in efflux of nystatin"
                     /note="putative transporter (ABC family)"
                     /codon_start=1
                     /transl_table=11
                     /product="NysH"
                     /protein_id="AAF71764.1"
                     /db_xref="GI:8050838"
                     /translation="MLLRLLRAQLRPYAWATAALVALQLVQILGTLLLPTLGAALIDQ
                     GVVRGDGGRITELGVVMGVVALVQIAAALGAAALAARTATAMGRDLRSALFRRILDFS
                     AREIGRFGTPSLLTRSVNDVQQVQNLAQTGFGIVVCAPLMCLGSVLLALRQDVPLALL
                     LVALVLVVAVCFGLLLARMGTLYARMQLTLDRLGRLLREAITGVRVVRSFVRDDHERA
                     RFAQTNDAFLVVSRRVGRLIATMLPVVLLLMNGFTVALLWTGSHRIDAGRMPIGSLSA
                     LLSYLSLILMSVVMLAFVFLSVPRARVCAGRIAEVLDTGSSVAPPAAPQPVRGPAGRI
                     ELCAAGYRYPGAEEPVLRDVDLTVEPGERIAVLGSTGSGKTTLLNLVLRLADATEGAV
                     RVGGTDVRELTAATLAAAVGFVPQRPYLFSGTVASNLRFGRPDATDEELWEALRVAQA
                     ADFVARMPDGLDAEITQGGGNVSGGQRQRLSLARALLRRPEIYLFDDCFSALDQATDA
                     ALRTALVPYTAGATVITVAQRISAGRDADRIVVLDRGRVVAQGTHDVLLRTSPTYREI
                     ALSQLTEEEAAHGLAGRP"
gene                 4714. .5748
                     /gene="nysDIII"
CDS                  4714. .5748
                     /gene="nysDIII"
                     /function="presumably involved in mycosamine biosynthesis"
                     /note="putative dGDP-mannose-4,6-dehydratase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysDIII"
                     /protein_id="AAF71765.1"
                     /db_xref="GI:8050839"
                     /translation="MSKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKSRV
                     SRLASELDFIDGDLMDQGSLVSAVDTVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                     VLRMLEAIRMVSGLSTSRTVSPRGQIRFYQASSSEMFGKAAETPQRETTLFHPRSPYG
                     AAKAYGHYITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKQGLQDKL
                     ALGNLDAVRDWGYAGDYVRAMHLMLQQDAGDDYVIGTGQMHSVRDAVRIAFEHVGLNW
                     EDYVVIDPDLVRPAEVEVLCADSAKAQDRLGWKPDVDFPTLMRMMVDSDLAQVSRENQ
                     YGDVLLAANW"
gene                 5930. .34363
                     /gene="nysI"
CDS                  5930. .34363
                     /gene="nysI"
                     /function="responsible for condensation steps 9 to 14 in
                     the nystatin polyketide backbone synthesis"
                     /note="polyketide synthase"
                     /codon_start=1
                     /transl_table=11
                     /product="NysI"
                     /protein_id="AAF71766.1"
                     /db_xref="GI:8050840"
                     /translation="MDNEQKLRDYLKLATADLRRTRRRVHKLESAAQEPVAIIGMTCR
                     YPGGVRSPEDLWRMVEAGEHGVTPFPTDRGWDLEALAAAPTASGGFLHDAPDFDADFF
                     GISPREAVAMDPQQRVVLESAWEAFERAGIDPTSVKGSRTGVFIGAMAQDYRVGPADG
                     AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                     AGGVTIMSGPGTFIEMGRQGGLSADGRCRSFGDTADGTGWAEGVGILVLERLSDAVRN
                     GHEILAVVRGTAVNQDGASNGLTAPNGPSQQQVIQQALVNARLAAGDIDVVEAHGTGT
                     TLGDPVEAQALLATYGQNRPADRPLLLGSVKSNLSHTQAAAGVAGVIKMVMAMRHGTL
                     PRTLHAEEPTHHVDWSQGAVRLLTDTTDWPATGAPRRAAVSSFGISGTNAHTIIEQAP
                     EPQPEDAATAQDDAAGSTPATAPVVPGVVPVLLSGRTPDALRGQAAALRAALDTGRRP
                     DLLDLAHSLATTRAGFEHRAVLLATDHPALTDGLTALADADDPAAAPAWITGTTRAET
                     RLAVLFTGQGAQRLGAGRELAARFPAFATALDAALDAFTPHLDRPLREVLWGTDAALL
                     DRTAYAQPALFAVEVALYRLIESFGVRPDHLAGHSVGEIVAAHLAGVLSLADAATLVA
                     ARGRLMQALPDGGAMIAVQASEADVAPLLAGHEDQVAIAAVNGPSAVVLSGAEATVTA
                     LAEQLAADGRKTRRLRVSHAFHSPLMEPMLDAFRAVVEDLTLQPPLLPVVSNLTGKPA
                     TVAQLTSADYWVDHVRHAVRFADGIDWLARHDTTAFLELGPDGVLSAMAQDCLDAADA
                     DAVTLPALRAGRPEEHTLTTALAGLHVHGATLDWTGCFAGTGARRTDLPTYAFQRRRY
                     WPKALQSGTADLRSVGLGAAHHPLLSAAVSLADAGGTLLTGRLSRQTHPWLADHTVRG
                     TTLLPGTAFLELAVRAGDEVGCDRVEELTLAAPLLLPEQGGVQVQLWIGNPDVSGRRT
                     VNVHARPDTGDDTPWTAHATGVLTTADASRQLPASSEQGGTPLAGDPHPALDAAQWPP
                     AGAEPLPLDGHYDRLADGGFGYGPVFQGLRAAWRGGDVVYAEVELPEAGRSDAEAFGL
                     HPALLDAALHAAPFTGLGERGRGGLPFSWEGVSLHAGGATTLRVRLTPVADDALALTV
                     ADGTGAPVLSVDSLVLRSVATQQLDTAAAVARDALFRLDWTPVQPTATDPGPVALLGA
                     DPFGLLTHAGFADAPAYPDLAALAAADGPVPTTVVLSLAGTGDDAADPARSAHRCAAE
                     ALAAVQTWLDHHERFAAARLVFVTRGATVGRDVAAAAVWGLVRSAQSENPGCFALVDL
```

```
DPDGAVGAAALVAALVSGEPQLAVRGDVLRVARLVRRPLTEVGAGADGTGDGVGDGSG
VSFSGEGAVLVTGGTGGLGAVLARHLVAEYGVRDLLLVSRSGERAVGAGELVAELAGV
GARVRVVACDVTDRAAVVELVGGHAVSAVVHAAGVLDDGMVGALTGERLSAVLRPKVD
AVWHLHEATRGLDLDAFVVFSSLAGVFGSPGQANYAAANAFLDALMTRRRAEGLPGLS
LAWGPWEQSGGMTGTLTDVDAERLARSGVPPLSVAQGLALFDAAVAGTDATCVPVRLD
LPVLRARGEVPPLLRSLIRVRARRAAVAGSATAGNLAQRLRRLDEDGRDEMVLDLVRG
QVALVLGHATGGDVDAGRAFRDLGFDSLTAVELRNRLNTVTGLRLPATLVFDYPTVRH
LATYVLDELLGTDAEVATVQPAAVAVADDPIVIVGMACRYPGGVSSPEDLWRVLTEGT
DAVSGFPTNRGWDVESLYHPDPDHPGTSYTRSGGFLHEAGEFDPGFFGMSPREALATD
SQQRLLLESSWEAIERAGIDPVSLRGSRTGVFAGVMYSDYSAMLASPEFEGFQGSGSS
PSLASGRVAYTLGLEGPAVTVDTACSSSLVAMHWAMQALRSGECGLALAGGVTVMSTP
AVFVDFARQRGLSPDGRCKAFADAADGVGWSEGVGVLVLERQSDAVRNGHEILAVVRG
SAVNQDGASNGLTAPNGPSQQRVIRQALASGGLTAGDVDVVEAHGTGTTLGDPIEAQA
LLATYGRDREPERPLLLGSVKSNLGHTQAAAGVAGVIKMVLAMRHGVVPRTLHVDAPS
SHVDWSEGAVELLSEQAAWPETGRVRRAGVSSFGISGTNVHVIVEQAPGAKAIAAAGA
ARRTPGAVPVLLSGRGRSALRGQAARLLGHLQARPDAELVDVALSLATTRSRFEQRAA
VVAQDRDQLIASLGALAADRPDPAVVEGEAAGRGRTAVLFTGQGSQRAAMGRELHEVQ
PEFAAAFDAVCAVFDPLLDRPLREVVFAEDGSDEAALLDETGWTQPALFAVEVALFRL
VESWGVRPDFVAGHSIGEIAAAHVAGVLTLEDACRLVAARATLMQALPTGGAMIAIQA
TEDEIAAHLDDTVAIAAVNGPQSVVISGDEEAAETIAATFAERGRKTKRLRVSHAFHS
PRMDGMLDAFRIVAEGLTYRAPRIPLVSDLTGRRADDAEVCTAEYWVRHVREAVRFAD
CVRTLRDAGATTFLELGSDGLLTAMAEDTLGDDHDAELVPMLRAGRAEELAAATALAR
LQVRGVDVDWAAYLAGTGARRTDLPTYAFQHAYYWPQLPTPAAALAAADPADQQLWAA
VERGDARELADILGLGEQDLTPLDSLLPALTSWRRGNQEKHLLDTLRYRVEWTRLSKP
TAPVLDGTWLLVASDATAADQPALLDGLADALGSHGARVRRLLLDDSCADRAVLAERL
ARTADVDAATQVLSVLPLDERDADDCPPLTRGLALTVALVQALADTGAQGRLWTATRG
AVSTNPADPVTHPVQAAAWGLGRGVALEHPRLWGGLVDLPQVFDERAGQRLAGILAVK
DAPDGEDQVALRATGVSGRRLVRHTVEALPTAAEFTATGTVLITGGTGGLGAEVARWL
ARAGAQHLVLTSRRGPDAPGAAELRAELEGYGPSVSVVACDVADRDALAAVLTALPEE
LPLTGVVHTAGVGHYGPLDTLSTAEFAGLTAAKLAGAAHLDALLADRELDFFVLFGSI
AGVWGSGNQSAYGAANAYLDALALHRRARGLAATSVAWGPWAEAGMAADDAVSETLRR
QGLGLLDPAPAMTELRRAVVRQDVTVTVADVDWQRYAPLFTSARPSALIAGLPEVRAL
AADERTEQDATGASEVVTRVRALAEPEQLRLLTDLVRTESATVLGHSSADAVPEGRAF
RDVGFDSLTAVELRKRLGAATGLSLPSTMVFDYPTPLELAQYLRAEILGAVLEVAGPV
ATGGADDEPIAIIGMACRFPGGVSSPEQLWDLVASGTDAISEFPVNRGWQTGHLFDPD
PDRPGTTYSTQGGFLHEADEFDPTFFGISPREALVMDPQQRLLLETTWESFERAGIRP
ETLRSTLTGTFVGSSYQEYGLGAGDGTEGHMVTGSSPSVLSGRLSYVFGLEGPAVTVD
TACSSSLVALHLACQSLRNGESNLAVAGGATIMTTPNPFIAFSRQRALAKDGRCKAFS
DDADGMTLAEGVGVVLVERLSDAQRNGHPVLAVLRGSAINQDGASNGLTAPNGPSQQR
VIRQALANARLAPGDIDALEAHGTGTPLGDPIEAQALFATYGRDRDPESALLLGSVKS
NIGHTQSAAGIASVIKMVMALRHSELPPTLHADAPSSHVDWSAGTVRLLTQARAWPET
GRPRRAAVSSFGISGTNAHVLLEQAPVADTPAEERPAVAPVPIAAGVVPWVVTARSAA
ALRGQAERLLAHAETVGTALPAAGPLDIGLSLVSARARFEHRAVVVPPAGTDPLAALR
AVATDGPSPVVARGVADVEGRTVFVFPGQGSQWVGMGSQLLDESAVFAERIAECAAAL
AEFTDWSLVDVLRGVVGAPSLERVDVVQPASFAVMVSLAALWRSRGVLPDAVVGHSQG
EIAAAVVSGALSLRDGARVVALRSQAIGRALAGRGGMMSVALSVDVLEPRLVEFEGRV
SVAAVNGPRSVVVAGEPEALDALHARLTADDIRARRIAVDYASHSHQVEDLHEELLEV
LAELAPRTSEVPFFSTVTGDWLDTARMDAGYWFRNLRGRVRFADAVADLLAAEYRAFV
EVSSHPVLSMAVQEAIDEAGVPAVAAGTLRRDQGGTDRFLLSAAEVFVRGVDVDWAGL
FEGTGASRIDLPTYAFQHEHLWAVPPAPEAVAAADPDDAAFWTAVEDGDVSALTAALG
TDEDSVAAVLPALTSWRRARRDRSTVDAWRYRVAWKPLGGTLPHPSLTGTWLLVTADG
IDDTDVAGALETYGAEVRRLVLDEECVDRAVLRERLAGAEDVTGIVSVLAAAERTDAV
PGTSLVLGTALTVALIQALGDAEIDAPVWALTRGAVSTGRADELTAPVQAQVTGIGWT
AALEHPQRWGGTLDLPAALDARAAQRLAAVLSGALGSDDQLAIRPSGVFTRRIVRAEA
```

```
Query Match           48.2%;  Score 257.2;  DB 1;  Length 123580;
Best Local Similarity   67.6%;  Pred. No. 6e-29;
Matches  361;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||||||||||||||||| |||||||  |||||||  | ||||||  |||||||||
Db      85772 GGAACGGTGTTGGTGACGGGTGGTACGGGTGGTCTGGGTGGGTTGGTGGCCCGGCATCTG 85831

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |   ||||||||||||||| || || |||||   ||  ||    ||||  || |||| ||
Db      85832 GTGCGGTCGTGTGGGGTGCGGCATTTGTTGTTGACCAGTCGTTCTGGTGTGGGTGCTGCG 85891

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              |||||||  ||  ||   ||||||  |||   |||||||  |||   |||   |  |||  |||
Db      85892 GGTGCGGCCGGGTTGGTCGCGGAGTTGGAGTCGTTGGGCGCGCGGGTTGTGGTTGCGGCG 85951

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||| ||  |   ||  || |   | || | |  ||||  | |   || |
```

```
Db      85952 TGTGATGTGGGTGATGGCTCGGCTGTTGCGGAGTTGGTTGCCGGTGTGTCGGAGTCGTAT 86011

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||| || | |  || |||||||||||||||||| |||| || |    | | |  ||  |
Db      86012 CCGTTGTCTGCGGTGGTGCATGCGGCTGGTGTGTTGGATGACGGTGTGGTGGGTTCGTTG 86071

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||||| |||||| |||   |||||||   | ||||||||||  ||        ||| ||
Db      86072 ACGCCGGAGCGGTTGGCTGCGGTGTTGCGTCCGAAGGTGGATGGTGCGTGGAACCTGCAT 86131

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||  |||||| ||| |||| |||    ||||| ||  | |||||  || |  |||||  |
Db      86132 GAGGCGACGCGTGGTCTGGATCTGGACGCGTTTGTTGTCTTCTCGTCTGTTGCGGGTGTG 86191

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | |||   || || |||| |||||| || ||||  |||||   | ||||||| ||  |
Db      86192 TTCGGGGGTGCGGGTCAGGCCAACTATGCGGCGGGTAATGCGTTTTTGGACGCGTTGATG 86251

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                | |||  ||| |  ||  |||| || ||||||  ||||||||||    |||
Db      86252 GTTCATCGGGTGGCTGGTGGGTTGCCTGGTGTGTCGTTGGCGTGGGGTGCTTGG 86305
```

```
RESULT 11
AX211739
LOCUS       AX211739              125401 bp    DNA     linear   PAT 06-SEP-2001
DEFINITION  Sequence 35 from Patent WO0159126.
ACCESSION   AX211739
VERSION     AX211739.1  GI:15523950
KEYWORDS    .
SOURCE      Streptomyces noursei.
  ORGANISM  Streptomyces noursei
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 125401)
  AUTHORS   Zotchev,S.B., Sekurova,O.N., Fjaervik,E., Brautaset,T.,
            Stroem,A.R., Valla,S., Ellingsen,T.E., Sletta,H.V. and
            Gulliksen,O.M.
  TITLE     Gene cluster encoding a nystatin polyketide synthase and its
            manipulation and utility
  JOURNAL   Patent: WO 0159126-A 35 16-AUG-2001;
            Norges Teknisk Naturvitenskapelige Universitet (NO) ; STIFTELSEN
            IND OG TEKNISK FORSKNING VED NORGES TEKNISKE HOGSKOLE (NO) ;
            ALPHARMA AS (NO) ; Sinvent AS (NO) ; Zotchev, Sergey Borisovich
            (NO) ; Sekurova, Olga Nikolayivna (NO) ; Fjaervik, Espen (NO) ;
            Brautaset, Trygve (NO) ; Strom, Arne Reidar (NO) ; Valla, Svein
            (NO)
FEATURES             Location/Qualifiers
     source          1. .125401
                     /organism="Streptomyces noursei"
                     /db_xref="taxon:1971"
                     /note="ATCC 11455"
BASE COUNT    15664 a  49692 c  42871 g  17174 t
ORIGIN
```

```
Query Match           48.2%;  Score 257.2;  DB 6;  Length 125401;
Best Local Similarity   67.6%;  Pred. No. 6e-29;
Matches  361;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||||||||||||||||| |||||||  |||||||  | ||||||  |||||||||
Db      86180 GGAACGGTGTTGGTGACGGGTGGTACGGGTGGTCTGGGTGGGTTGGTGGCCCGGCATCTG 86239

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |   ||||||||||||||| || || |||||   ||  ||    ||||  || |||| ||
Db      86240 GTGCGGTCGTGTGGGGTGCGGCATTTGTTGTTGACCAGTCGTTCTGGTGTGGGTGCTGCG 86299

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              |||||||  ||  ||   ||||||  |||   |||||||  |||   |||   |  |||  |||
Db      86300 GGTGCGGCCGGGTTGGTCGCGGAGTTGGAGTCGTTGGGCGCGCGGGTTGTGGTTGCGGCG 86359

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||| ||  |   ||  || |   | || | |  ||||  | |   || |
```

```
Db      86360 TGTGATGTGGGTGATGGCTCGGCTGTTGCGGAGTTGGTTGCCGGTGTGTCGGAGTCGTAT 86419

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300

Db      86420 CCGTTGTCTGCGGTGGTGCATGCGGCTGGTGTGTTGGATGACGGTGTGGTGGGTTCGTTG 86479

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360

Db      86480 ACGCCGGAGCGGTTGGCTGCGGTGTTGCGTCCGAAGGTGGATGGTGCGTGGAACCTGCAT 86539

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420

Db      86540 GAGGCGACGCGTGGTCTGGATCTGGACGCGTTTGTTGTCTTCTCGTCTGTTGCGGGTGTG 86599

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480

Db      86600 TTCGGGGGTGCGGGTCAGGCCAACTATGCGGCGGGTAATGCGTTTTTGGACGCGTTGATG 86659

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534

Db      86660 GTTCATCGGGTGGCTGGTGGGTTGCCTGGTGTGTCGTTGGCGTGGGGTGCTTGG 86713


RESULT 12
AF357202
LOCUS       AF357202               113193 bp    DNA     linear   BCT 17-JUL-2001
DEFINITION  Streptomyces nodosus amphotericin biosynthetic gene cluster,
            complete sequence.
ACCESSION   AF357202
VERSION     AF357202.1  GI:14794889
KEYWORDS    .
SOURCE      Streptomyces nodosus.
  ORGANISM  Streptomyces nodosus
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 113193)
  AUTHORS   Caffrey,P., Lynch,S.V., Flood,E.M., Finnan,S.M. and Oliynyk,M.
  TITLE     The amphotericin biosynthetic gene cluster from Streptomyces
            nodosus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 113193)
  AUTHORS   Caffrey,P., Lynch,S.V., Flood,E.M., Finnan,S.M. and Oliynyk,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2001) Industrial Microbiology, University College
            Dublin, Belfield, Dublin, Ireland
FEATURES             Location/Qualifiers
     source          1. .113193
                     /organism="Streptomyces nodosus"
                     /db_xref="taxon:40318"
     gene            complement(4. .1824)
                     /gene="amphG"
     CDS             complement(4. .1824)
                     /gene="amphG"
                     /note="possible amphotericin export protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphG"
                     /protein_id="AAK73498.1"
                     /db_xref="GI:14794890"
                     /translation="MAPDVPEEHEEERESEQPVRRLAALLRPHRRSVGLALTAGVVGI
                     LLNAFGPLLLGRVTDLIADGVLGHGGPAPGVDFGALGRLLMILLVLYVVASVFMLVQN
                     WLVASVVRLLIHDLRHRAQEKLARLPLRYFDRKPAGETLSRGTDDVDNLQQTLQQTLT
                     DLISSVFSLVIMLSLMLIISPSLAGVMLLSVPVSGLLAAWISKRAQPQYAAQWSASGK
                     LTAHVEEMCAGHALVKAFDRRAEAEQRFDERNEAVYRAGSGAQFASGAIEPVMMFVAN
                     LGYVAVAVVGAWKVVNGSLTLGDVQAFILYARQFSQPIVEIASVAGRLQSGVASAQRV
                     FTLLDAPEQEPEPDRPLAVERVEGRVEFQDVSFRYSPDTPLIEGLSLSVEPGSTVAVV
                     GPSGAGKTTVANLLMRFYEIDSGRILLDGTDTAAMNRDDLRSRFGLVLQDTWLFKGTI
                     AENIAYGSPGATRADIVEAARATYADRFIRTLSQGYDTVLDDESGGVSAGEKQLITVA
                     RAFLARPAVLVLDEATSSVDTRTELLIQRAMNTLRPGRTSFVIAHRLSTIRDADVIVV
                     MKSGRIVEQGTHDQLIDAQGAYARLHAARADAPAADVTVG"
     gene            complement(1805. .3628)
                     /gene="amphH"
     CDS             complement(1805. .3628)
                     /gene="amphH"
                     /note="possible amphotericin export protein"
                     /codon_start=1
                     /transl_table=11
                     /product="AmphH"
                     /protein_id="AAK73499.1"
                     /db_xref="GI:14794891"
                     /translation="MAPSVRWPTPFQRAAAGRRQPGCHSVLLRLMRTQLRPYAGSVVA
                     LVVLHLVQILGTLLLPTLGAALIDEGVVRHDSDRIGTIGATMAVVALVQIAAALGAAA
                     LGARTSTALGRDLRSAVFRRVLDFSAREIGRFGTPSLLTRSVNDVQQVQNLAQSGLGI
                     FVAAPLMCLGSVLLALRQDVTLALILVPMVLVVAVCFGLLLSRMAALYARLQQTLDRI
                     GRLLRERITGVRVVRSFARDAHEGERFTRTNEELLGLSLGVGRLIAVMLPSVLLLMNL
                     FTLGLLWVGARRIDSGSMQIGALSAFLSYLSLILMSVVMLAFVFLNVPRARVCAERIT
                     EVLQAETDVVPPASPRPMAGPAGQVELVGAEFRYPGAENAVLRDLSLTLRPGERVAVL
                     GSTGSGKTTLLHLILRLVDVTAGEVRIGGTDVRELDPSVMAAAVGYVPQRPYLFAGTV
                     ASNMRFGRPDATDEELWEVLRIAQHEGFVTRLGGLDTEIAQGGTTVSGGQRQRLAIAR
                     ALLRRPAIYLFDDSFSALDQSTEAALRKALVPYTEGATVITVAQRVASVRDADRIVLL
                     DQGGIAATGTHDALLRDSPTYREIALSQRTREETAHGAGRS"
gene            3840. .4874
                /gene="amphDIII"
CDS             3840. .4874
                /gene="amphDIII"
                /note="mycosamine biosynthesis protein"
                /codon_start=1
                /transl_table=11
                /product="AmphDIII"
                /protein_id="AAK73500.1"
                /db_xref="GI:14794892"
                /translation="MPKRALITGITGQDGSYLAEHLLSQGYQVWGLIRGQANPRKFRV
                SRLASELSFVDGDLMDQGSLVSAVDKVQPDEVYNLGAISFVPMSWQQAELVTEVNGMG
                VLRVLEAIRMVSGLSMSRTAGTEGQIRFYQASSSEMFGKVAETPQRETTLFHPRSPYG
                AAKAYGHFITRNYRESFGMYAVSGMLFNHESPRRGQEFVTRKISLAVARIKLGLQDKL
                ALGNMDAVRDWGYAGDYVRAMHLMLQQDAPDDYVIGTGEMHTVRDAVRFAFEHVGLDW
                KDYVVVDPDLVRPAEVEVLCADSSKAQAQLGWKPSVDFQELMRMMVDADLASVSRQNE
                LDDLLLAHSW"
gene            5042. .33574
                /gene="amphI"
CDS             5042. .33574
                /gene="amphI"
                /note="polyketide synthase multienzyme polypeptide housing
                extension modules 9, 10, 11, 12, 13, and 14"
                /codon_start=1
                /transl_table=11
                /product="AmphI"
                /protein_id="AAK73501.1"
                /db_xref="GI:14794893"
                /translation="MDNEQKLRDYLKLATADLRRARRRVGELESASQEPIAIVGMTCR
                YPGGVSSPEDLWRMVEAGENGVTPFPTDRGWDLEALASSPTSRGGFLHDAPEFDADFF
                GISPREAVAMDPQQRVVLESAWEAFERAGINPTSVKGSRTGVFIGAMAQDYRVGPADG
                AEGFQLTGNTGSVLSGRISYTFGTVGPAVTVDTACSSSLVAVHLATQALRAGECTLAL
                AGGVTVMSGPGTFIEMGRQGGLSVDGRCRSFGDTADGTGWAEGVGILVLERLSDAIRN
                GREILAVVRGTAVNQDGASNGLTAPNGPSQQAVIEQALVNARLSAGDIDVVEAHGTGT
                TLGDPVEAQALLATYGQQRDEDKPLLLGSVKSNISHTQAAAGVAGVIKMVMAMRHGVL
                PRTLLADEPTRHVDWSQGAVRVLTENTEWPATGAPRRAAVSSFGISGTNAHTIVEQAP
                EPEPADPEDDAPSTPAAVTGVLPVLLSGRSPEVLRAQAAALLTTLGTGTPPAPADLAY
                SLATTRTAFEHRAVLLASDLPELTGRLTAIAEGTDPAVLADTVTGTARTETRLAVLFT
                GQGAQRLGAGRELAARFPVFAAALDAALDAFTPHLDVPLRKVLWGEDADRLDRTEYAQ
                PALFAVEVALYRLLESFEVKPDHLAGHSVGEIAAAHVAGVFSLDDAATLVAARGRLMQ
                ALPEGGAMVAVQASEDEVAPLLAGHEDLVSLAAVNGPSAVVLSGDETTVTELAARLAA
                DGRKTSRLRVSHAFHSPLMAPMLDEFRNVVEGLTLHSPLLPVVSDVTGEPATVAQLTS
                PDYWVDHVRQAVRFADGIDWLARHDVTAFLELGPDSVLSAMAQNCLDAAGSDALTVPA
                LREGRPEDHTFTAALAALHTQGTALHWDACFTGTGARRTDLPTYTFQRRRYWPRAVQG
                GAADLRSVGLGAAHHPLLSAAVSLADSEGALLTGRISLLSHPWLADHTVRGATLLPGT
                AFLELAVRAGDEVGCDRVDELTLAAPLVLPEQGGVQVQLWIGNPDASGRRSVTVYGRP
                DADEDAPWTSHATGVLSASRTTSDFDATVWPPADAETLPVDGLYERLAEGGFGYGPLF
                QGLRAAWRRGDEVFAEVVLPESGHTDAESFGLHPALLDSALHAASFVDLDERAAGGLP
                FSWEGVSLHASGATTLRVRLAPAAGDAVAIAVADDSGQLVLSADSLILRAVAAREIDA
                AAALVRDALFRLDWVPVTAVAASGTAAALVGEDPFGLRALPQFGDLAVHPDLADLAAA
                DGAVPDTVLLPLTGTGPDADPVTAAHRAATEALAAVRTWLEQDERFAASRLALVTRGA
                TTGHDPAAAAVWGLVRSAQSENPGRFLLVDLDADQDTPALPAAALTSEEPQLAVRGEE
                LRAARLVRRPASTAEAVPAFGGEGAVLVTGGTGGLGAVLARHLVAEHGVRELVLVSRR
                GGAAAGAAELVAELAESGARATVVACDVTDRAAVAELVAAHPVSAVVHSAGVLDDGMV
                GTLTPERLTTVLRPKVDAAWNLHEATRDLDLKAFVLFSSVAGVLGSPGQANYAAGNAF
```

```
LDALAAHRRAAGLPGLSLAWGPWEQTGGMTGGISEDDLRRMARAGTPALTVEQGLALL
DAALDGDDAALAPVRLDLSVLRAQGEVPPLLRSLIRGRSRRAAVAGSATAGGLAQRLA
RLDAESRDELVLDLVRGQVALVLGHATGAEIDAGRAFRELGFDSLTAVELRNRLNTVT
GLRLPATLVFDYPTVSHLASYVLDELLGTEVEAEVVQRGTAAVADDPIVIVGMACRYP
GGVTSPEDLWRLVTEGTDAVSGFPVNRGWDVENLYHPDPDHPGTAYTRSGGFLHEAGE
FDPGFFGMSPREALATDSQQRLLLEASWEAIERAGIDPVGLRGSATGVFAGVMYSDYS
AMLGSPEFEGFQGSGSSPSLASGRVSYTLGLEGPAVTVDTACSSSLVAMHWAMQALRS
GEISLALAGGVTVMSTPAVFVDFARQRGLSPDGRCKAFSDSADGVGWSEGVGMLVLER
QSDAIRNGHQILAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALASGGLSAGDVDVV
EAHGTGTTLGDPIEAQALLATYGRDRDPEQPLLLGSVKSNIGHTQAAAGVAGVIKMVM
SMRHGVLPRTLHVDAPSSHVDWTEGAVELLTEQTAWPETGRPRRAAVSSFGISGTNVH
TVLEQAPGTTVPAPAAPERTAGAVPLLLSGRTRDALRAQAARLLTHLQNHPEPSLADL
GHSLATTRSRFERRAAVIAQDREGLLASLGSLAAGRPDPAVVEGEAAGRARVAVMFSG
QGSQRAAMGRELYETQPRFAAAPDEVCAALDPLLDRPLREVVFAAEGSEEAALLDRTG
WTQPALFAVEVALYRLVESWGVRADFVTGHSIGEIAAAHIAGVFTLQDAARLVAARAT
LMEALPSGGAMVAVQATEEEVAPLLGEGLSVAAVNGPTSVVVSGDEDPAVELAAEFSG
RGRRTKRLRVSHAFHSPHMDAMLDAFRTVAETLSYAAPRIPLVSDLTGRRADDAEVRT
ADYWVRHVREAVRFADCVRTLRDAGATLFLELGPDGLLTAMAEDTLGDERYDHNTALV
PLLRADRPEESAAATAAARLQIHGVDLDWTAYLAGTGARRVDLPTYAFQHAHYWPQLP
SAAPSPAGDPADQKLWAAVERGDAAELAAVLGLDEDSLTPLDSLLPALSSWRRGNQEK
ALLDTLRYRVEWTRLSKPAAPVLDGTWLLVSSDATADDETELLDGLAEALGAHGARVR
RLVLDADCADRAVLGARLADTENADNTAQVLSVLPLDERPTDGPAGFTQGLALTIALV
QALADTGAHGRLWTATRGAVSTGPADPVTHPAQATAWGMGRGVALEHPRLWGGLVDLP
ADFDRGAGQRLAEVLAVKDAPDGEDQVALRATGVHGRRLVRHIVDELPSADQFTASGS
VLITGGTGGLGAETARWLARSGAAHLVLTSRRGPDAPGAAELRAELEQSGASVSIVAC
DVADRDALAAVLDGLSADQPLTGVVHTAGVGHYGPLDALTPAEFAGLTAAKLAGAAHL
DNLLGDRELDFFILFGSIAGVWGSGDQSAYGAANAYLDALALARRARGLAATSIAWGP
WGGTGMAADDAVSGTLRRQGLGLLDPAPALTEMRRAVVRQDVTVTVADVDWTRYAPLF
TSARPSALISDLPEVRALAAENTPADTGDASEIVQRVRSLSEPEQLRLLTDLVRTEAA
TVLGHSSAGAVPEDRAFREIGFDSLTAVELRKHLGAATGLSLPSTMVFDYPTPLELAQ
YLRAEMVGSVLEVAGPVATGGTDDEPIAIIGMSCRYPGGVSSPEQLWDLVLSGTDAIT
DFPVNRGWNTAGLYDPDPDHPGTTYSTQGGFLHEADEFDPMFFGISPREALVMDPQQR
LLLETTWEAFERAGLTPDTLRGSLTGTFIGSSYQEYGMGAGDGAEGHLVTGTSPSVLS
GRLAYVFGLEGPAVTVDTACSSSLVALHLACQALRNGESNLAVAGGATVMTTPNAFVA
FSRQRALAQDGRCKAFSESADGMTLAEGVGIVLVERLSDARRNGHPVLAVIRGSAINQ
DGASNGLSAPNGPSQQRVIRQALANARVAPGEIDLLEAHGTGTPLGDPIEAQALFATY
GRTRTPETALLLGSVKSNIGHSQSAAGVASIIKMVMALRHGVMPQTLHADEPSSHVDW
SPGTVRLLGENTDWPQTGRPRRAAVSSFGISGTNAHVILEQETEAPAAEDEQLAPAPL
PVAAGVVPWLLSARGAAALREQADRLLTHLVTADPAARPIDIGLSLATSRALFEHRAV
VVPPAGTDPLEALRAVAADGPSGVVARGVADVAGRTVFVFPGQGSQWAGMGAQLLDES
PVFAERIAECAAALAEFTDWNLIDVLRGAEGAPTLERVDVVQPASFAVMVSLAAVWRA
QGVEPDAVVGHSQGEIAAAVVSGALSLRDGARVVTLRAQAIGRSLAGRGGMMSVALPV
AEVEARLEAFEGRVSVAAENGPRSSVVAGEPEALDELHAQLTAEEIRARRVAVDYASH
SPHVEDLHDEILELLAEVAPRTSEIPFFSTVTGDWLDTTVMDAGYWYRSLRGRVLFAD
AVRDLIAADHRAFIEVSSHPVLAMSVQDMIDDAGVAGVASGTLRRDNGGLDRFLLSAA
EVFVRGVQVDWAAVFEGTGASRVDLPTYAFQHENLWAMAAAPEAVTAADPEDAAFWTA
VEDGDVSALTAALGTDEDSVAAVLPALSSWRRARKERSTVDSWRYRPTWKPVTKLPQR
TLDGTWLLVSADGVDDTDVAEALETGGAEVRRLVLDESCTDRAVLRERLTDADGLTGI
VSVLAGAERTGAVPGTGLVLGVALTVALVQALGDAGIDTPLWALTRGAVSTGRSDKVT
APVQAQVTGIGWTAALECPGRWGGVVDLPETLDARAGQRLAAVLAGALGDDDQIALRS
SGVFTRRIVRADAAPDGSARDWKPRGTTLVTGGSGTLAPHLARWLAEQGAEHLVLVSR
RGPEAPGAAELRAELAERGTETTLAACDITDRDAVAALLESLKAEGRTVRTVVHTAAT
IELHTLDATTLDDFDRVLAAKVTGAQILDELLDDEELDDFVLYSSTAGMWGSGAHAAY
VAGNAYLAALAEHRRARGLTALSLSWGIWADDLQLGRVDPQMIRRSGLEFMDPQLALS
GLKRALDDDEQVIAVADVDWETYHPVYTSARPTPLFDEVPEVQRLTAAAEQSAGDPAR
GEFAAALLALPAAEQHRKLLETVRTEAASVLGLSSAEDLTDQRAFRDVGFDSLTAVGL
RNRLASVTGLTLPSTMVFDYPNPAALAGFLHSELADVHSAGAVAVTAGAPVDDDPIAI
VGMSCRYPGGITSAEQLWRVSLEEVDAVSVFPADRGWDAEALYDPDPDASGRTYSVQG
GFLRDVADFDPGFFGISPREALSMDPQQRLLLETAWEVFENAGLDPVAQRGSRTGTFI
GASYQDYGAAVPGSEGSEGHMITGSLPSVLSGRVSYLFGLEGPAVTLDTACSSSLVAI
HLACQSLRNGESTLALAGGASIMSTPMSFIGFSRQRALAEDGRCKAYAEGADGMTLAE
GVGLILLERLSDARRNGHEVLAVIRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAG
VEANDIDVLEGHGTGTALGDPIEAQALFATYGKDRDPERPVLLGSVKSNIGHTQMASG
VASIIKLVHALREGVAPKSLHIDQPSTHVDWSSGTIQLLTERTEWPETGRPRRAAVSS
FGLSGTNVHTVLEQAPAADAPAAEDTPAPRDALVPVLVSGRGEAALRAQAGALLDLLA
ERPGIHPTDLAFSLATSRAALEHRAAVVADDHEALVRGLTALRDGLPGAGLVQGRTGR
GRTAFLFTGQGSQRLGMGRELYERHPVFADALDAVLARIDGTTERPLRDVLFAAEGSQ
DAALLHRTGYAQPALFALEVALFRLLESWGVTPDYLAGHSVGEIAAAHVAGVLDLDDA
CTLVAARGRLMQALPEGGAMVALEAAEDEVLPHLEGLADQVSVAAVNGPRSVVVAGEE
EPVLALAAHFAEQGRRTKRLRVSHAFHSPLMDPMLDDFAAVARALTYHAPSIPFVSNV
TGTLAAPEQVCTADYWVSHVRSAVRFADGIGWLSTQGGVQTFLELGPDGVLSGMARES
LTDASRTALLPTLRGDRPEEQALVTAVAAAHAHGFDVDWTAWFQGSGARRVALPTYAF
QRERYWPDTTAAGITAPAPGSALDAEFWAAVEHADVASLTASLGLDDATVTAMVPALT
```

```
                                                AWRQRRGEQSALDSWRYRVTWKPRGGAPGAAPTGRWLVLVPAEHRDEATAAWAADVEA
                                                ALATATVRVEVTGTDRAALAARLTEAADGDTFQGVLSLLALAPGDAGHPGAPAALTLT
                                                ATALQALGDARIDAPLWNITRGAVAVGRSEQVTAPEQAAVWGLFRAAALELPARVGGS
                                                VDLPEDLDTQAARRLRGILAAADGEDAVAVRASGVFLRRLAHSPAADTVGSAFDPAAG
                                                TVLITGGTGGIGGHLARRLARDGAAHLLLTSRRGPDAPGAGELRAELEESGARVTIAA
                                                CDAADRDALAALLATVPEDAPLTAVFHTAGVVDDHVVDELTPESFATVLHAKTVAARH

  Query Match             47.2%;  Score 252;  DB 1;  Length 113193;
  Best Local Similarity   68.7%;  Pred. No. 3.5e-28;
  Matches  367;  Conservative    0;  Mismatches  155;  Indels   12;  Gaps    1;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  |||||||||||||||||||| || ||||  |||||||||| | ||||||| ||||||
Db       9245 GGTGCGGTGTTGGTGACGGGTGGTACCGGTGGTCTGGGTGCGGTGCTGGCGCGTCATCTG 9304

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  | | |  || ||||| || ||| |  |||| || || |||||||   |  ||  ||
Db       9305 GTGGCCGAGCACGGTGTGCGTGAGCTGGTTCTGGTCAGTCGTCGTGGTGGTGCGGCGGCG 9364

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              |||||||  |  |||  |||||||||||| | || ||| |||  || |  | | || ||
Db       9365 GGTGCGGCCGAGCTGGTCGCGGAGCTGGCGGAGTCGGGTGCGCGGGCGACGGTGGTGGCC 9424

Qy        181 TGTGATGTGGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              |||||||||  ||| ||   || ||||| |  |||| ||                |||| | |
Db       9425 TGTGATGTGACGGACCGTGCGGCGGTGGCCGAGCTGGTG-----------GCCGCGCAT 9472

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||| || |||  || ||||| | || ||||| |||||||| |  | | ||| | | ||
Db       9473 CCGGTGTCGGCGGTGGTGCACTCCGCCGGTGTCCTGGACGACGGCATGGTCGGCACGCTG 9532

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              || ||||| || ||   ||| ||| ||    |||||||||| || ||      ||  ||
Db       9533 ACCCCCGAACGCCTCACCACCGTGCTCCGTCCCAAGGTGGACGCGGCCTGGAACCTCCAT 9592

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||  || || |  |||| |||  ||| || ||||||||||| |||| ||| ||  ||
Db       9593 GAGGCCACCCGCGACCTGGACCTGAAGGCCTTTGTGCTGTTCTCGTCGGTCGCCGGCGTC 9652

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              |||||     |||| |||| ||||||||||| | ||| |||  |||||||| || |||
Db       9653 CTGGGCAGCCCCGGTCAGGCCAACTACGCCGCAGGCAACGCCTTCCTGGACGCCCTCGCG 9712

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                ||  || ||||| ||||| |||||||  |||||||||| ||||| |  |||
Db       9713 GCCCACCGCCGGGCCGCGGGCCTGCCGGGTCTGTCGCTGGCCTGGGGCCCCTGG 9766


RESULT 13
SFU78289
LOCUS       SFU78289               43280 bp    DNA     linear   BCT 13-AUG-1997
DEFINITION  Streptomyces fradiae tylactone synthase, starter module and modules
            1-7, (tylG) gene, complete cds.
ACCESSION   U78289
VERSION     U78289.1  GI:2317859
KEYWORDS    .
SOURCE      Streptomyces fradiae.
  ORGANISM  Streptomyces fradiae
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 43280)
  AUTHORS   DeHoff,B.S., Sutton,K.L. and Rosteck,P.R. Jr.
  TITLE     Sequence of Streptomyces fradiae tylactone synthase gene tylG
  JOURNAL   Unpublished (1996)
REFERENCE   2  (bases 1 to 43280)
  AUTHORS   DeHoff,B.S., Sutton,K.L. and Rosteck,P.R. Jr.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-NOV-1996) Eli Lilly and Company, Lilly Corporate
            Center, Indianapolis, IN 46285, USA
COMMENT     tylG.
FEATURES             Location/Qualifiers
     source          1. .43280
                     /organism="Streptomyces fradiae"
```

```
                /db_xref="taxon:1906"
gene            1. .43280
                /gene="tylG"
CDS             816. .14234
                /gene="tylG"
                /note="ORF1; putative"
                /codon_start=1
                /transl_table=11
                /product="tylactone synthase starter module and modules 1
                & 2"
                /protein_id="AAB66504.1"
                /db_xref="GI:2317860"
                /translation="MSSALRRAVQSNCGYGDLMTSNTAAQNTGDQEDVDGPDSTHGGE
                IAVVGMSCRLPGAAGVEEFWELLRSGRGMPTRQDDGTWRAALEDHAGFDAGFFGMNAR
                QAAATDPQHRLMLELGWEALEDAGIVPGDLTGTDTGVFAGVASDDYAVLTRRSAVSAG
                GYTATGLHRALAANRLSHFLGLRGPSLVVDSAQSASLVAVQLACESLRRGETSLAVAG
                GVNLILTEESTTVMERMGALSPDGRCHTFDARANGYVRGEGGGAVVLKPLDAALADGD
                RVYCVIKGGAVNNDGGGASLTTPDREAQEAVLRQAYRRAGVSTGAVRYVELHGTGTRA
                GDPVEAAALGAVLGAGADSGRSTPLAVGSVKTNVGHLEGAAGIVGLIKATLCVRKGEL
                VPSLNFSTPNPDIPLDDLRLRVQTERQEWNEEDDRPRVAGVSSFGMGGTNVHLVIAEA
                PAAAGSSGAGGSGAGSGAGISAVSGVVPVVVSGRSRVVVREAAGRLAEVVEAGGVGLA
                DVAVTMADRSRFGYRAVVLARGEAELAGRLRALAGGDPDAGVVTGAVLDGGVVVGAAP
                GGAGAAGGAGAAGGAGGGGVVLVFPGQGTQWVGMGAGLLGSSEVFAASMRECARALSV
                HVGWDLLEVVSGGAGLERVDVVQPVTWAVMVSLARYWQAMGVDVAAVVGHSQGEIAAA
                TVAGALSLEDAAAVVALRAGLIGRYLAGRGAMAAVPLPAGEVEAGLAKWPGVEVAAVN
                GPASTVVSGDRRAVAGYVAVCQAEGVQARLIPVDYASHSRHVEDLKGELERVLSGIRP
                RSPRVPVCSTVAGEQPGEPVFDAGYWFRNLRNRVEFSAVVGGLLEEGHRRFIEVSAHP
                VLVHAIEQTAEAADRSVHATGTLRRQDDSPHRLLTSTAEAWAHGATLTWDPALPPGHL
                TTLPTYPFNHHHYWLDTIDGGGGDDATQEKESGPLTRELRGLPSSQKQLGFLLDLVCR
                HTAVVLGLDTAAEVDPDLSFKKQGIQSMTGVELRNRLLTETGLALPTTLVYDRPTPRA
                LAQFLHTELLDGSPSGSVLAPAQKSFEAQEPIAVVGMGCRFPGGVGSPEALWRLVVEG
                VDAVSPFPGDRGWDVEGLYDPEPGVAGKSYVREGGFLHDAAEFDAEFFGISPREAVAM
                DPQQRLLLETSWEAIERAGIDPHSLHGSRTGVYAGVMPQEYGPRLAEGAEGSDGYLLT
                GTSGSVVSGRVAYTLGLEGPAVTVDTACSSSLVALHLAVQALRGGECDMALAGGVTVM
                AGPGMFVEFSRQRGLAADGRCKAFADGADGTAWAEGAGVVLVERLSDARRLGHPVLAV
                VCGSAVNQDGASNGLTAPSGPSQERVIRQALGNARLTVADVDVVEAHGTGTRLGDPIE
                AQALLGTYGRDRDGGRPVWLGSLKSNIGHAQAAAGVAGVIKMVLAMRYGWLPRTLHVD
                EPSRHVDWSAGGVWLLTEAREWPGVDRPRRAAVSAFGVSGTNAHLILEAPDTAEAESA
                TTPVRSEVSESAAVLDARSGVVPVVVSGRSRVVVREAAGRLAEVVEAGGVGLADVAVT
                MAGRSRFGYRAVVLARGEAELAGRLRALAGGDPDAGVVTGAVVDPETGSGGGGVVLVF
                PGQGTQWVGMGAGLLGSSEVFAASMRECARALSVHVEWDLLEVVSGGAGLERVDVVQP
                VTWAVMVSLARYWQAMGVDVAAVVGHSQGEIAAATVAGALSLEDAAAVVALRAGLIGR
                YLAGRGAMAAVPLPAGEVEAGLAKWPGVEVAAVNGPASTVVSGDRRAVAGYVAVCQAE
                GVQARLIPVDYASHSRHVEDLKGELERVLSGIRPRSPRVPVCSTVAGEQPGEPVFDAG
                YWFRNLRNRVEFSAVVGGLLEEGHRRFIEVSAHPVLVHAIEQTAEAADRSVHATGTLR
                RQDDSPHRLLTSTAEAWAHGATLTWDPALPPGHLTTLPTYPFNHHHYWLDTTPTTPAT
                TTQSPTDAWRYRVTWKALTESSPVRPHSIGRCLLVAPPTTDGELLDGLTTVLSERGAS
                VARLEVPIGARRAEVAELLKPSMESAGEENTTVVSLLGLVPSTDAVRTSIALLQAVSD
                IGVPAARVWALTRRAVAVVPGETPQDAGAQLWGFGRVAALELPDIWGGLIDLPETAEL
                TRTPETSQPPQTPERLPQTPNRRALELAAAVLAGRDGEDQVAVRASGIYGRRVSRAAA
                AGAASWQPSGTVLITGGMGAIGRRLARRLAAEGAERLVLTSRRGPEAPGAAELAEELR
                GHGCEVVHAACDVAERDALAALVTAYPPNAVFHTAGILDDAVIDTLSPESFETVRGAK
                VCGAELLHQLTADIKGLDAFVLFSSVTGTWGNAGQGAYAAANAALDALAERRRAAGLP
                ATSVAWGLWGGGGMAAGAGEESLSRRGLRAMDPDAAVDALLGAMGRNDVCVTVVDVDW
                ERFAPATNAIRPGRLFDTVPEAREALTAAGTTSATPDGAPELARRLSMLNETERLRKL
                VELVRTEAAFVLRHPNTDAIGAERPFKSAGFDSLTSLELRNRLNAGTGLKLPATVIFD
                HPSPTALARLLLDRLTGAGAPAPAADEPPLPVAVADDDPVVIVGMACRFPGGAGTPEA
                LWKLVTEERDVIGAAPTDRGWDLDSVYDPEPGVAGKTYVREGGFLHDAAEFDAEFFGI
                SPREAVAMDPQQRLLLETSWEAIERAGIDPHSLHGSRTGVYVGLTHQEYASRLHEAPE
                EYEGYLLTGKSASVVSGRISYTLGLEGPSLSIDTACSSSLVALHNAAQALRGGECDMA
                LAGGVTVMAAPGLFVEFSRQRGLAADGRCKAFADGADGTAWAEGAGVVLVERLSDARR
                LGHPVLAVVCGSAVNQDGASNGLTAPSGPSQERVIRQALANARLTVADVDVVEAHGTG
                TRLGDPIEAQALLGTYGRDRDAECPVWLGSLKSNIGHAQAAAGVAGVIKMVLAMRYGW
                LPRTLHVDEPSRHVDWSAGGVRLLTEAREWPGVDRPRRAAVSAFGVSGTNAHLILEAP
                EALEALEATDAPEAPEAPEAPDVTDVTEALEAPDATEAEGAKAPGSPEEAQPAVGVVP
                VVVSGRSRVVVREAAGRLAEVVEAGGVGLADVAVTMAGRSRFGYRAVVLARGEAELAG
                RLRALAGGDPDAGVVTGAVVDPETGSGGGGVVLVFPGQGTQWVGMGAGLLGSSEVFAA
                SMRECARALSVHVEWDLLEVVSGGAGLERVDVVQPVTWAVMVSLARYWQAMGVDVAAV
                VGHSQGEIAAATVAGALSLEDAAAVVALRAGLIGRYLAGRGAMAAVPLPAGEVEAGLA
                KWPGVQVAAVNGPASTVVSGDRRAVAGYVAVCQAEGVQARLIPVDYASHSRHVEDLKG
                ELERVLSGIRPRSPRVPVCSTVAGEQPGEPVFDAGYWFRNLRNRVEFSAVVGGLLEQG
                HRRFIEVSAHPVLVHAIEQTAEAADRSVHATGTLRRQDDSPHRLLTSTAEAWAHGATL
                TWDPALPPGHLTTLPTYPFNHHHYWAVTSPAGVGDAAAGRFGMTWEDHPFLRGGLPLA
                DSGERVFAGRLAGSEHDWLTDHAVSGVTLLPGTAFVEFALHAGAATGCGRLEELSVEA
                PLVLPAAGGVRVQMRVSAADESGRRRVAIHSAPEAAVHSAAEGGDSAGVWTRHGEGTL
                VPDPEPTPPDADWARAWPPAGERVEPAELYERFGALGYEYGEAFAGVRAVWRQPDALL
                AEVLLPDRASTGAGRFGVHPALLDAALQPWIAGGLLEVPEDAVLLPFAWQGVSLYATG
                AGALRVRLTKAGDGAVSLQAADTSGAAVLSLGALVMRPLARRKLDVLLGTDAGERSLY
                RVEWQPRLLPAGPPRSWAVLGPDADRLAGTPGLGDQPDGGPTALYPEVRALRKALAAG
                APRPEAVVLPVLSGAGATPESVRQTTERCLTALQDWLDAEELVDTPLIVLTRGAVAAV
                PGEEIGDLACAGVWGLVRSARSEHPGRFALVDTDGHPDDRTALPLALRAVLDGAGQLS
                LRAGTARTPVLLRAGTPEEQRGPAFDPAGTVLVTGATGTLGRLLARHLAAEHGVRHLL
                LLSRGGRAAEGADELAAELAGLEAEPCFAACDAADREALARVLAEVPADRPLTGVIHA
                AGVLDDGTLDALTPERIGTVMRPKADAALNLHELTRTSPLSVFAVFSGAAGILGRPGQ
                ANYAAANTFLDALAQHRRAHGLPAVSLAWGLWGGATGMTGHLSGTDLRRMRRSGIAPM
                THDQGLALFDRALAASAEDPLLVPMRLDLAALVRERAEHGPDAVPGPLLGLLPARAAV
                RQAAAPVRGGAPAPAGGEGTAERLAGLGEEARLRELVRLVRAEVSGVLGYSGPDAVEP
                GRPFKDLGFDSLTAVELRNRLGAATGLRLPTALVFDRPTSQAVAEYLAAELAGPRDGG
                DTAAAAFEGLEALAAAVGALAEDDLRRDVLRRRLTELAAALTPQGRNPSAPAPAPSDL
                DERLDSANDDDLFAFIEEQL"
misc_feature    942. .3929
                /gene="tylG"
                /function="encodes starter module"
misc_feature    942. .2156
                /gene="tylG"
                /function="encodes beta-ketoacyl ACP synthase KS-Q of
                starter module"
misc_feature    2571. .3557
                /gene="tylG"
                /function="encodes acyltransferase domain of starter
                module"
misc_feature    3675. .3929
                /gene="tylG"
                /function="encodes acyl carrier protein domain of starter
                module"
misc_feature    3993. .8471
                /gene="tylG"
                /function="encodes module 1"
misc_feature    3993. .5264
                /gene="tylG"
                /function="encodes beta-ketoacyl ACP synthase of module 1"
misc_feature    5631. .6617
                /gene="tylG"
                /function="encodes acyltransferase domain of module 1"
misc_feature    7410. .7949
                /gene="tylG"
                /function="encodes beta-ketoreductase domain of module 1"
misc_feature    8220. .8471
                /gene="tylG"
                /function="encodes acyl carrier protein domain of module
                1"
misc_feature    8541. .13970
                /gene="tylG"
                /function="encodes module 2"
misc_feature    8541. .9812
                /gene="tylG"
                /function="encodes beta-ketoacyl ACP synthase of module 2"
misc_feature    10260. .11246
                /gene="tylG"
                /function="encodes acyltransferase domain of module 2"
misc_feature    11319. .11876
                /gene="tylG"
                /function="encodes dehydratase domain of module 2"
misc_feature    12861. .13415
                /gene="tylG"
                /function="encodes beta-ketoreductase domain of module 2"
misc_feature    13719. .13970
                /gene="tylG"
                /function="encodes acyl carrier protein domain of module
                2"
CDS             14351. .19945
                /gene="tylG"
                /note="ORF2"
                /codon_start=1
                /transl_table=11
                /product="tylactone synthase module 3"
```

```
                     /protein_id="AAB66505.1"
                     /db_xref="GI:2317861"
                     /translation="MTAELVATRKRLGALEERAREPIAVVAMSCRYPGGVTTPEDLWR
                     LLADERDAVSGLPRDRGWDLDALYDPDGGPGTSYAREGGFLSHCAGFDAEFFGISPRE
                     ALAMDPQQRLLLETSWEALERAGVTADRARGSRTGVYAGVMYDDYGARVLYGAGAGPP
                     EDLEGYLVNGSAGSIASGRVSYTFGLRGPAVTVNTACSSSLVSLHLAVRALRNGECDM
                     ALAGGATVLSTPTVLVDFSRQRGLAPDGRCKAFADSADGTSWAEGAGMLLLQRLSDAR
                     REGRPVLAVIRGSAVNQDGASNGLTAPNGRAQRQVIEDALRDAGVGPDQVDAVEAHGT
                     GTELGDPIEAGALLATYGTARTAERPLWLGSLKSNIGHTQAAAGVAGVIKMVLAMRHG
                     RLPRTLHVDRPTTRVDWEKGGVRLLTEPVPWPGEAGEPRRAGVSSFGASGTNAHVVLE
                     SVPAGEPPAAGRPEDTGGAWTVSGRGPAALRAQAARLYDALTGTGTGTGQGAGQGAGP
                     GTAEVAGALAHARTAFRHRAVVLGGNRAELLAGLRELAEEEHPGPRVVTGTAPATERR
                     TAFLFSGQGSQRAGSGRGLYRRHPVFARALDEVCAALEPHLHRPLRDLMFAEPGSPEA
                     EPLDRTEFTQPALFALQTALFRLAEHHGLRAEALCGHSVGEIAAAHAAGVLTLPDAAR
                     LVAARGRLMQALPAGGAMAALRATAEEIAPLLERRAGELALAAVNGPSSVVVSGDEAA
                     VLELLEQWRAEGREARRLAVSHAFHSPRMDGMLTQFDRVARTLTFAPPTIPLVSTLTG
                     TPVTEETLCTADHWVRQAREPVRFLDAMRTLRADGIDTFVELGPDGVLSAMARDCADD
                     RPDGDTTGAGDGETPDPLLTLPLLRRSVPETGDAEHPGGFERALATAYAHGVPLRLAP
                     APDAASLAVAAELPTYAFQRTHYWLDAPAAPAALPAGLDDAGHPLLSAALDLPGGRGT
                     VWTGALSAATLPWAADHSVHGRTVLPGTALLDLALHAAPRVGELTFEAPLVLPEDGEV
                     RLRVVLAEPDASGVRELSVHSAGEDGGWTRHATAVLDTGTTTAGEPAGAPPAAWPPGD
                     AEPLDLAAEYERFADAGIGYGPAFRGLRSAWRDGDAILADVRLPGELAGEADRYGIHP
                     ALLDAALHTAAAALGGAHGMLPFTWNGVTLHARGAHAIRVRLTPAGPDAVAVTAVDPA
                     GRPVFTAASLTLRPVTTGQLTAAEAARAPLYRVRWTGLPDTGTARDHTWAVAGGPGDL
                     LPGETPHHPDLASALADTGTAPFRVLADLRGYGTATPRELASQALALVQQWADAAEAA
                     EGRLVLVTRRAVDIGDGVTDPAAATVWGLVRAAQSEHPGCFALLDTDDSPRSRQLLPR
                     VAGTAEQLALRDGTLLAPSLTRATLPAGARLPALDGTVLITGGTGSLGAEAARHLVTR
                     HGARRLLLTSRSGPQAPGAAELVAELAALGAHADVAACDVADRAALRALLDRVPAGHP
                     LTAVLHTAGVLDDGVLTAQTPQRLAAVLRPKADAVRNLHELTQGHALSAFILYSSAAG
                     VLGSAGQSGYAAANAYLDSFAVWRRSRGLPAVSLGWGPWDGGGMASGLGGTDTARLRR
                     SGIAPLSRAEGLAALDAALAAGGDDTAPAHLLPIRVDAVTLRGADTVPAVLRDLAGTA

  Query Match           47.0%;  Score 250.8;  DB 1;  Length 43280;
  Best Local Similarity  66.9%;  Pred. No. 6.4e-28;
  Matches  357;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||  ||||||||||| | ||| ||   ||| ||   |  | |||| ||||||||||
Db      12864 GGCACGGTCCTGGTGACGGGCGCGACCGGCACGCTCGGGCGGCTGCTGGCCCGGCATCTG 12923

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              || |  | |   || |||||  |||||||| || |||||||  | || | || |||
Db      12924 GCCGCCGAGCACGGTGTGCGCCATCTGCTGCTGCTGAGCCGCGGCGGCCGGGCTGCCGAA 12983

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || ||||| |  ||    |||||| ||||||| |||||  || |||  | |  | |  ||
Db      12984 GGCGCGGACGAACTCGCCGCGGAACTGGCCGGGTTGGAAGCCGAGCCGTGCTTCGCGGCC 13043

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              |||||  | ||  |||  ||    ||   |||  |||  ||||| ||  ||  |||||    |
Db      13044 TGTGACGCGGCGGACCGCGAGGCCCTGGCACGGGTGCTGGCGGAGGTGCCGGCCGACCGG 13103

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||  ||  ||   | ||  ||||| ||  |||||  |||||  |   ||  |||  | | ||
Db      13104 CCGCTGACCGGAGTGATCCACGCGGCCGGGGTGCTCGACGACGGCACACTCGACGCGCTG 13163

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||  ||  ||  |||  |  ||  ||  ||   |  | |  ||||  |||  ||  ||  ||     ||| |
Db      13164 ACCCCGGAACGGATCGGTACCGTCATGCGGCCGAAGGCGGACGCGGCGCTGAACCTGCAC 13223

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||  |||||  |||    |    ||||||||  |||||  | |  |||||     ||  ||  ||  |||
Db      13224 GAACTGACCCGGACCAGCCCGCTGTCGGTGTTCGCGGTCTTCTCGGGCGCTGCCGGCATC 13283

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              |||||     ||||  |||| ||||||||||||| |||||   ||    ||  ||||||||  |||
Db      13284 CTGGGCCGCCCCGGACAGGCCAACTACGCCGCCGCCAACACCTTCCTCGACGCGCTCGCG 13343

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                | |  ||  ||  ||    ||  ||  ||  |  ||||||||||| ||||||||||||
Db      13344 CAGCACCGCCGCGCCCACGGCCTCCCCGCCGTGTCGCTGGCCTGGGGGCTGTGG 13397
```

```
RESULT 14
MTV018/c
LOCUS       MTV018                 53450 bp    DNA     linear   BCT 03-AUG-2001
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 90/162.
ACCESSION   AL021899 AL123456
VERSION     AL021899.1  GI:3242282
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis H37Rv.
  ORGANISM  Mycobacterium tuberculosis H37Rv
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 53450)
  AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrell,B.G.
  TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
  JOURNAL   Nature 393 (6685), 537-544 (1998)
  MEDLINE   98295987
   PUBMED   9634230
REFERENCE   2  (bases 1 to 53450)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 20, 1998 this sequence version replaced gi:2896750.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES             Location/Qualifiers
     source          1. .53450
                     /organism="Mycobacterium tuberculosis H37Rv"
                     /strain="H37Rv"
                     /db_xref="taxon:83332"
     source          51045. .>53450
                     /organism="Mycobacterium tuberculosis H37Rv"
                     /strain="H37Rv"
                     /db_xref="taxon:83332"
                     /clone="Y63A"
     misc_feature    <1. .51145
                     /note="fragment designated v018. Does not represent a
                     physical clone"
     gene            complement(114. .1370)
                     /gene="Rv2015c"
     CDS             complement(114. .1370)
                     /gene="Rv2015c"
                     /note="Rv2015c, (MTV018.02c), len: 418. Nearly identical
                     to Mycobacterium tuberculosis Rv1765c(MTCY28.31c, 378 aa),
                     an orf starting next to ISB9, and ending in IS6110.
                     Different N-terminus chosen and C-terminus differs as that
```

of MTCY28.31c has been truncated by IS6110. Does NOTshow similarities with transposes. BLAST hits with non-IS part of MTU78639. FASTA scores:. Z95890|MTCY28_31 (378 aa) opt: 2417 z-score: 2553.1 E(): 0; 97.8% identity in 364 aa overlap. TBparse score is 0.939"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2015c"
/protein_id="CAA17229.1"
/db_xref="GI:2896752"
/db_xref="SPTREMBL:O53461"
/translation="MSSTATSGAAVVSPAERVEVLFEELAELAGQRNAIDGRIVEIVA
ELDRDGLWGVTGARSVAGLVAWKMGCSSGNAHTIATVARRLPEFPRCARGMREGRLSL
DQVGVIAGRAGEGSDAHYAQLAGVATVNQLRTALKLEPRPEPEPDFRPEPRPSITRSA
DEQFSCWRIKLPHVEAAKFDAALQSHLDALIAEYKRDHDNSDGVSDQRPPLPGNVEAF
LRLVEAGWDAEVARRPHGQHTTVVMHLDVQERAAGLHLGPLLSESERRYLLCDATFEA
WFERDGQVIGCGRTTRQINRRLRRALEHRDRTCVVPGCGATRGLHAHHIRHWQDGGAT
ELANLVLVCPYHHRAHHRGLITITGPADNLTVADSAGRPLSAGSLARASTKPPPAVAP
WPGPTGERADWWWYEPFQPQPPPISN"
RBS 1715. .1719
/note="possible RBS, ggcga, for Rv2016"
gene 1724. .2299
/gene="Rv2016"
CDS 1724. .2299
/gene="Rv2016"
/note="Rv2016, (MTV018.03), len: 191. Unknown. TBparse score is 0.927"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2016"
/protein_id="CAA17230.1"
/db_xref="GI:2896753"
/db_xref="SPTREMBL:O53462"
/translation="MTELGDKFLAALVGTIRDTRFDIADMRNWRPGWFPTMHSRCLSN
LIHDRIWAHLVTLIASNPGTSIKDKGATREIVVGAHLRLRIKRHHAGDEISTYPTRTA
IEFWQQGSQPAFPGLEEVRIAVGYRWDPDTREIGAPLLSLRDGKDHVIWVVELDEPAA
GVKITWTPIEPTLPSIDFGDLGEDSGASGER"
RBS 2287. .2292
/gene="Rv2016"
/note="possible RBS, ggggga, for Rv2017"
gene 2296. .3336
/gene="Rv2017"
CDS 2296. .3336
/gene="Rv2017"
/note="Rv2017, (MTV018.04), len: 346. Unknown but shows similarity at N-terminal end to several transcriptional regulators e.g. Bacillus subtilis BSUB0012_44 (108 aa). ContainsPS00142 Neutral zinc metallopeptidases, zinc-binding regionsignature in C-terminal half, may be fortuitous. .FASTA scores: Z99115|BSUB0012_44 Bacillus subtilis (108 aa) opt: 154 z-score: 199.9 E(): 0.0012; 35.5% identity in 62 aa overlap. TBparse score is 0.908. Contains probable helix-turn-helix motif at aa 18-39 (Score 2243, +6.83 SD)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2017"
/protein_id="CAA17231.1"
/db_xref="GI:2896754"
/db_xref="SPTREMBL:O53463"
/translation="MNGLGDVLAVARKARGLTQIELAELVGLTQPAINRYESGDRDPD
QHIVAKLAEILGVTDDLLIHGNRFRGALAVDAHMRRHKTTKASAWRQLEARLNLLRVH
ASFLFEEVAINSEQHVPAFDPEFTAAEDAARLVRAQWRMPMGPVVNLTRWMEAAGCLV
FEEDFATQRIDGLSQWVDDYPVMLINANAAPDRKRLTLAHELGHLVLHSTNPTENMET
EATAFAAEFLMPESEIRPELRRLDLGKLLELKREWGVSMQALLARAYRMGLVSAEART
KLYKAMNARGWKTKEPGIESIVREKPSLPAHIGMTLRSRGFTDQQAAAIAGYANPADN
PFRPEGGRLHAI"
misc_feature 2884. .2913
/gene="Rv2017"
/note="PS00142 Neutral zinc metallopeptidases, zinc-binding region signature"
RBS 3565. .3569
/note="possible RBS, ggagg, for Rv2018"
gene 3578. .4297
/gene="Rv2018"
CDS 3578. .4297
/gene="Rv2018"
/note="Rv2018, (MTV018.05), len: 239. Unknown but similar to MTCY339.01c (238 aa). FASTA scores: Z77163|MTCY339_1 Mycobacterium tuberculosis cosmid (238 aa) opt: 142 z-score: 175.1 E(): 0.029; 24.8% identity in 250 aa overlap. TBparse score is 0.929. Contains probable helix-turn-helix motif at aa 215-236 (Score 1175, +3.19 SD)"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2018"
/protein_id="CAA17232.1"
/db_xref="GI:2896755"
/db_xref="SPTREMBL:O53464"
/translation="MAGDQELELRFDVPLYTLAEASRYLVVPRATLATWADGYERRPA
NAPAVQGQPIITALPHPTGSHARLPFVGIAEAYVLNAFRRAGVPMQRIRPSLDWLIKN
VGPHALASQDLCTDGAEVLWRFAERSGEGSPDDLVVRGLIVPRSGQYVFKEIVEHYLQ
QISFADDNLASMIRLPQYGDANVVLDPRRGYGQPVFDGSGVRVADVLGPLRAGATFQA
VADDYGVTPDQLRDALDAIAA"
gene 4287. .4703
/gene="Rv2019"
CDS 4287. .4703
/gene="Rv2019"
/note="Rv2019, (MTV018.06), len: 138. Unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2019"
/protein_id="CAA17233.1"
/db_xref="GI:2896756"
/db_xref="SPTREMBL:O53465"
/translation="MQPDRNLLADLDHIFVDRSLGAVQVPQLLRDAGFRLTTMREHYG
ETQAQSVSDHKWIAMTAECGWIGFHKDANIRRNAVERRTVLDTGARLFCVPRADILAE
QVAARYIASLAAIARAARFPGPFIYTVHPSKIVRVL"
gene complement(4719. .5018)
/gene="Rv2020c"
CDS complement(4719. .5018)
/gene="Rv2020c"
/note="Rv2020c, (MTV018.07c), len: 99. Unknown. TBparse score is 0.891"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv2020c"
/protein_id="CAA17234.1"
/db_xref="GI:2896757"
/db_xref="SPTREMBL:O53466"
/translation="MAPGMKWAAKTDHLAIVLLPRHHRRHSRRGRALPARSRSALGWI
IERYRVTTDKASGIVNDPNDWCDEHDDPTYIVDLIKKVTTVSVETMKIVDGLAGG"
gene complement(5103. .5408)
/gene="Rv2021c"
CDS complement(5103. .5408)
/gene="Rv2021c"
/note="Rv2021c, (MTV018.08c), len: 101. Unknown, possible regulatory protein similar to M. tuberculosis

Query Match 47.0%; Score 250.8; DB 1; Length 53450;
Best Local Similarity 66.9%; Pred. No. 6.1e-28;
Matches 357; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

```
Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  |||||||| | || ||||| || ||  ||  |||| ||||||||||| || ||  ||
Db      34157 GGCACGGTGTTGATCACCGGTGGCACCGGGATGGCGGGTTCGGCGGTGGCCCGTCACGTG 34098

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
                | | |     |||||||||  ||||| ||||||||||||||||| || ||||||||||
Db      34097 GTGGCTCGTCATGGGGTGCGCAATCTGGTGTTGGTGAGCCGGCGCGGCCCGGATGCTCCC 34038

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || || | ||  |||  ||| ||| |||||||     || ||| ||||| | | || ||
Db      34037 GGGGCTGCGGAGCTGGTGGCCGAGTTGGCCGCCGCCGGTGCGCAGGTGCAGGTGGTGGCC 33978
```

```
Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||| | || ||| ||   || | ||| |  | || | |  | | |||| |    |  |
Db      33977 TGTGACGCGGCGGATCGAGCGGCGTTGGCCAAGGTGATCGCCGATATTCCGGTTCAGCAT 33918

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||  || |||| ||  | ||  | || || |  || ||||| |||  | |    || ||
Db      33917 CCATTGTCGGGCGTGATCCACACCGCCGGCGCACTCGACGACGCGGTGGTGATGTCACTG 33858

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              || || || ||| |||    ||||||   | ||||||||| || ||        ||| |
Db      33857 ACACCGGATCGGGTGGATGTGGTGTTGCGGTCCAAGGTGGACGCGGCGTGGCACCTGCAC 33798

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||| |||| || |   |||| |||||||||| || |||| || |||   || ||| |
Db      33797 GAGTTGACTCGCGACCTGGATGTGTCGGCGTTTGTCATGTTTTCGTCGATGGCCGGGCTG 33738

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | || ||| | || |||| |||||| || || |||||  |   | |||| |||||||
Db      33737 GTCGGATCGTCGGGCCAGGCCAACTATGCGGCCGCCAATTCGTTTTTGGATGCGCTGGCC 33678

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                ||  ||||||||    || |||||||   | || ||||  ||||| ||||||
Db      33677 GCCCACCGGCGGGCCCATGGGCTGCCGGCCATCTCCCTGGGCTGGGGTCTGTGG 33624
```

```
RESULT 15
AE007061/c
LOCUS       AE007061               15970 bp    DNA     linear   BCT 27-APR-2001
DEFINITION  Mycobacterium tuberculosis CDC1551,  section 147 of 280 of the
            complete genome.
ACCESSION   AE007061 AE000516
VERSION     AE007061.1  GI:13881778
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis CDC1551.
  ORGANISM  Mycobacterium tuberculosis CDC1551
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE   1  (bases 1 to 15970)
  AUTHORS   Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
            Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
            Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
            Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
            Gill,J., Mikula,A. and Bishai,W.
  TITLE     Whole genome comparison of Mycobacterium tuberculosis clinical and
            laboratory strains
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 15970)
  AUTHORS   Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
            Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
            Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
            Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouri,H.,
            Gill,J., Mikula,A. and Bishai,W.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-APR-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .15970
                     /organism="Mycobacterium tuberculosis CDC1551"
                     /strain="CDC1551"
                     /db_xref="taxon:83331"
                     /note="clinical strain"
     gene            70. .726
                     /gene="MT2106"
     CDS             70. .726
                     /gene="MT2106"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK46385.1"
                     /db_xref="GI:13881779"
                     /translation="MRIAALVAVSLLIAGCSREVGGDVGQSQTIAPPAPAPSAAPSTP
                     PAAGAPITTIVSWIEAGHPVDPAAYHVATRDGVTTQLGDDVAFSASSGTVACMTDARH
                     TSGTLACLVRLANPPPRPETAYGEWKGGWVDFDGIHLQVGSARADPGPFVYGNGPELA
                     NGDTLSIGDYRCRSYQAGLFCVNYAHQSAVRFASAGIEPFGCLKPAPPPDGVGVAFGC
                     "
     gene            complement(763. .3327)
                     /gene="MT2107"
     CDS             complement(763. .3327)
                     /gene="MT2107"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAK46386.1"
                     /db_xref="GI:13881780"
                     /translation="MRIAVTGASGVLGRGLTARLLSQGHEVVGIARHRPDSWPSSADF
                     IAADIRDATAVESAMTGADVVAHCAWVRGRNDHINIDGTANVLKAMAETGTGRIVFTS
                     SGHQPRVEQMLADCGLEWVAVRCALIFGRNVDNWVQRLFALPVLPAGYADRVVQVVHS
                     DDAQRLLVRALLDTVIDSGPVNLAAPGELTFRRIAAALGRPMVPIGSPVLRRVTSFAE
                     LELLHSAPLMDVTLLRDRWGFQPAWNAEECLEDFTLAVRGRIGLGKRTFSLPWRLANI
                     QDLPAVDSPADDGVAPRLAGPEGANGEFDTPIDPRFPTYLATNLSEALPGPFSPSSAS
                     VTVRGLRAGGVGIAERLRPSGVIQREIAMRTVAVFAHRLYGAITSAHFMAATVPFAKP
                     ATIVSNSGFFGPSMASLPIFGAQRPPSESSRARRWLRTLRNIGVFGVNLVGLSAGSPR
                     DTDAYVADVDRLERLAFDNLATHDDRRLLSLILLARDHVVHGWVLASGSFMLCAAFNV
                     LLRGLCGRDTAPAAGPELVSARSVEAVQRLVAAARRDPVVIRLLAEPGERLDKLAVEA
                     PEFHSAVLAELTLIGHRGPAEVEMAATSYADNPELLVRMVAKTLRAVPAPQPPTPVIP
                     LRAKPVALLAARQLRDREVRRDRMVRAIWVLRALLREYGRRLTEAGVFDTPDDVFYLL
                     VDEIDALPADVSGLVARRRAEQRRLAGIVPPTVFSGSWEPSPSSAAALAAGDTLRGVG
                     VCGGRVRGRVRIVRPETIDDLQPGEILVAEVTDVGYTAAFCYAAAVVTELGGPMSHAA
                     VVAREFGFPCVVDAQGATRFLPPGALVEVDGATGEIHVVELASEDGPALPGSDLSR"
     gene            complement(3332. .15787)
                     /gene="MT2108"
     CDS             complement(3332. .15787)
                     /gene="MT2108"
                     /note="similar to PID:987099; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="polyketide synthase"
                     /protein_id="AAK46387.1"
                     /db_xref="GI:13881781"
                     /translation="MVDQLQHATEALRKALVQVERLKRTNRALLERSSEPIAIVGMSC
                     RFPGGVDSPEGLWQMVADARDVMSEFPTDRGWDLAGLFDPDPDVRHKSYARTGGFVDG
                     VADFDPAFFGISPSEALAMDPQHRMLLELSWEALERAGIDPTGLRGSATGVFAGLIVG
                     GYGMLAEEIEGYRLTGMTSSVASGRVAYVLGLEGPAVSVDTACSSSLVALHMAVGSLR
                     SGECDLALAGGVTVNATPTVFVEFSRHRGLAPDGRCKPYAGRADGVGWSEGGGMLVLQ
                     RLSDARRLGHPVLAVVVGSAVNQDGASNGLTAPNGPSQQRVVRAALANAGLSAAEVDV
                     VEGHGTGTTLGDPIEAQALLATYGQDRGEPGEPLWLGSVKSNMGHTQAAAGVAGVIKM
                     VLAMRHELLPATLHVDVPSPHVDWSAGAVELLTAPRVWPAGARTRRAGVSSFGISGTN
                     AHVIIEAVPVVPRREAGWAGPVVPWVVSAKSESALRGQAARLAAYVRGDDGLDVADVG
                     WSLAGRSVFEHRAVVVGGDRDRLLAGLDELAGDQLGGSVVRGTATAAGKTVFVFPGQG
                     SQWLGMGIELLDTAPAFAQQIDACAEAFAEFVDWSLVDVLRGAPGAPGLDRVDVVQPV
                     LFAVMVSLAELWKSVAVHPDAVIGHSQGEIAAAYVAGALSLRDAARVVTLRSKLLAGL
                     AGPGGMVSIACGADQARDLLAPFGDRVSIAVVNGPSAVVVSGEVGALEELIAVCSTKE
                     LRTRRIEVDYASHSVEVEAIRGPLAEALSGIEPRSTRTVFFSTVTGNRLDTAGLDADY
                     WYRNVRQTVLFDQAVRNACEQGYRTFIESSPHPALITGVEETFAACTDGDSEAIVVPT
                     LGRGDGGLHRFLLSAASAFVAGVAVNWRGTLDGAGYVELPTYAFDKRRFWLSAEGSGA
                     DVSGLGLGASEHPLLGAVVDLPASGGVVLTGRLSPNVQPWLADHAVSDVVLFPGTGFV
                     ELAIRAGDEVGCSVLDELTLAAPLLLPATGSVAVQVVVDAGRDSNSRGVSIFSRADAQ
                     AGWLLHAEGILRPGSVEPGADLSVWPPAGAVTVDVADGYERLATRGYRYGPAFRGLTA
                     MWARGEEIFAEVRLPEAAGGVGGFGVHPALLDAVLHAVVIAGDPDELALPFAWQGVSL
                     HATGASAVRARIAPAGPSAVSVELADGLGLPVLSVASMVARPVTERQLLAAVSGSGPD
                     RLFEVIWSPASAATSPGPTPAYQIFESVAADQDPVAGSYVRSHQALAAVQSWLTDHES
                     GVLVVATRGAMALPREDVADLAGAAVWGLVRSAQTEHPGRIVLVDSDAATDDAAIAMA
                     LATGEPQVVLRGGQVYTARVRGSRAADAILVPPGDGPWRLGLGSAGTFENLRLEPVPN
                     ADAPLGPGQVRVAMRAIAANFRDIMITLGMFTHDALLGGEGAGVVVEVGPGVTEFSVG
                     DSVFGFFPDGSGTLVAGDVRLLLPMPADWSYAEAAAISAVFTTAYYAFIHLADVQPGQ
                     RVLIHAGTGGVGMAAVQLARHLGLEVFATASKGKWDTLRAMGFDDDHISDSRSLEFED
                     KFRAATGGRGFDVVLDSLAGEFVDASLRLVAPGGVFLEMGKTDIRDPGVIAQQYPGVR
                     YRAFDLFEPGRPRMHQYMLELATLFGDGVLRPLPVTTFDVRRAPAALRYLSQARHTGK
                     VVMLMPGSWAAGTVLITGGTGMAGSAVARHVVARHGVRNLVLVSRRGPDAPGAAELVA
                     ELAAAGAQVQVVACDAADRAALAKVIADIPVQHPLSGVIHTAGALDDAVVMSLTPDRV
                     DVVLRSKVDAAWHLHELTRDLDVSAFVMPSSMAGLVGSSGQANYAAANSFLDALAAHR
                     RAHGLPAISLGWGLWDQASAMTGGLDAADLARLGREGVLALSTAEALELFDTAMIVDE
```

```
                     PFLAPARIDLTALRAHAVAVPPMFSDLASAPTRRQVDDSVAAAKSKSALAHRLHGLPE
                     AEQHAVLLGLVRLHIATVLGNITPEAIDPDKAFQDLGFDSLTAVEMRNRLKSATGLSL
                     SPTLIFDYPTPNRLASYIRTELAGLPQEIKHTPAVRTTSEDPIAIVGMACRYPGGVNS
                     PDDMWDMLIQGRDVLSEFPADRGWDLAGLYNPDPDAAGACYTRTGGFVDGVGDFDPAF
                     FGVGPSEALAMDPQQRMLLELSWEALERAGIDPTGLRGSATGVFAGVMTQGYGMFAAE
                     PVEGFRLTGQLSSVASGRVAYVLGLEGPAVSVDTACSSSLVALHMAVGSLRSGECDLA
                     LAGGVTVNATPTVFVEFSRHRGLAPDGRCKPYAGRADGVGWSEGGGMLVLQRLSDARR
                     LGHPVLAVVVGSAVNQDGASNGLTAPNGPSQQRVVRAALANAGLSAAEVDVVEGHGTG
                     TTLGDPIEAQALLATYGQDRGEPGEPLWLGSVKSNMGHTQAAAGVAGVIKMVLAMRHE
                     LLPATLHVDVPSPHVDWSAGAVELLTAPRVWPAGARTRRAGVSSFGISGTNAHVIIEA
                     VPVVPRREAGWAGPVVPWVVSAKSESALRGQAARLAAYVRGDDGLDVADVGWSLAGRS
                     VFEHRAVVVGGDRDRLLAGLDELAGDQLGGSVVRGTATAAGKTVFVFPGQGSQWLGMG
                     MGLHAGYPVFAEAFNTVVGELDRHLLRPLREVMWGHDENLLNSTEFAQPALFAVEVAL
                     FRLLGSWGVRPDFVMGHSIGELSAAHVAGVLSLENAAVLVAARGRLMQALPAGGAMVA
                     VQAAEEEVRPLLSAEVDIAAVNGPASLVISGAQNAVAAVADQLRADGRRVHQLAVSHA
                     FHSPLMDPMIDEFAAVAAGIAIGRPTIGVISNVTGQLAGDDFGSAAYWRRHIRQAVRF
                     ADSVRFAQAAGGSRFLEVGPSGGLVASIEESLPDVAVTTMSALRKDRPEPATLTNAVA
                     QGFVTGMDLDWRAVVGEAQFVELPTYAFQRRRFWLSGDGVAADAAGLGLAASEHALLG
                     AVIDLPASGGVVLTGRLSPSVQGWLADHSVAGVTIFPGAGFVELAIRAGDEVGCGVVD
                     ELTLAAPLVLPASGSVAVQVVVNGPDESGVRGVSVYSRGDVGTGWVLHAEGALRAGSA
                     EPTADLAMWPPAGAVPVEVADGYQQLAERGYGYGPAFRGLTAMWRRGDEVFAEVALPA
                     DAGVSVTGFGVHPVLLDAALHAVVLSAESAERGQGSVLVPFSWQGVSLHAAGASAVRA
                     RIAPVGPSAVSIELADGLGLPVLSVASMLARPVTDQQLRAAVSSSGPDRLFEVTWSPQ
                     PSAAVEPLPVCAWGTTEDSAAVVFESVPLAGDVVAGVYAATSSVLDVLQSWLTRDGAG
                     VLVVMTRGAVALPGEDVTDLAGAAVWGLVRSAQTEHPGRIVLVDSDAPLDDSALAAVV
                     TTGEPQVLWRRGEVYTARVHGSRAVGGLLVPPSDRPWRLAMSTAGTFENLRLELIPDA
                     DAPLGPGQVRVAVSAIAANFRDVMIALGLYPDPDAVMGVEACGVVIETSLNKGSFAVG
                     DRVMGLFPEGTGTVASTDQRLLVKVPAGWSHTAAATTSVVFATAHYALVDLAAARSGQ
                     RVLIHAGTGGVGMAAVQLARHLGLEVFATASKGKWDTLRAMGFDDDHISDSRSLEFED
                     KFRAATGGRGFDVVLDSLAGEFVDASLRLVAPGGVFLEMGKTDIRDPGVIAQQYPGVR
                     YRAFDLFEAGPDRIAQILAELATLFGDGVLRPLPVTTFDVRCAPAALRYLSQARHTGK
                     VVMLMPGSWAAGTVLITGGTGMAGSAVARHVVARHGVRNLVLVSRRGPDAPGAAELVA
                     ELAAAGAQVQVVACDAADRAALAKVIADIPVQHPLSGVIHTAGALDDAVVMSLTPDRV
                     DVVLRSKVDAAWHLHELTRDLDVSAFVMFSSMAGLVGSSGQANYAAANSFLDALAAHR
                     RAHGLPAISLGWGLWDQASAMTGGLATVDFKRFARDGIVAMSSADALQLFDTAMIVDE
                     PFMLPAHIDFAALKVKFDGGTLPPMFVDLINAPTRRQVDDSLAAAKSKSALLQRLEGL
                     PEDEQHAVLLDLVRSHIATVLGSASPEAIDPDRAFQELGFDSLTAVEMRNRLKSATGL
                     ALSPTLIFDYPNSAALAGYMRRELLGSSPQDTSAVAAGEAELQRIVASIPVKRLRQAG
                     VLDLLLALANETETSGQDPALAPTAEQEIADMDLDDLVNAAFRNDDE"
BASE COUNT     3023 a   6045 c   4734 g   2168 t
ORIGIN

  Query Match           46.7%;  Score 249.2;  DB 1;  Length 15970;
  Best Local Similarity  66.7%;  Pred. No. 1.3e-27;
  Matches  356;  Conservative    0;  Mismatches  178;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  |||||||| | || ||||| || ||  ||  |||| |||||||||| || ||  ||
Db      10750 GGCACGGTGTTGATCACCGGTGGCACCGGGATGGCGGGTTCGGCGGTGGCCCGTCACGTG 10691

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  | |     |||||||||  ||||| ||||||||||||||||| || ||||||||||
Db      10690 GTGGCTCGTCATGGGGTGCGCAATCTGGTGTTGGTGAGCCGGCGCGGCCCGGATGCTCCC 10631

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || || | ||  |||  ||| ||| |||||||    || ||| |||| | | | || ||
Db      10630 GGGGCTGCGGAGCTGGTGGCCGAGTTGGCCGCCGCCGGTGCGCAGGTACAGGTGGTGGCC 10571

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||| | || ||| ||   || | ||| |  | || | |  | | |||| |    |  |
Db      10570 TGTGACGCGGCGGATCGAGCGGCGTTGGCCAAGGTGATCGCCGATATTCCGGTTCAGCAT 10511

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||  || |||| ||  | ||  | || || |  || ||||| |||  | |    || ||
Db      10510 CCATTGTCGGGCGTGATCCACACCGCCGGCGCACTCGACGACGCGGTGGTGATGTCACTG 10451

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              || || || ||| |||     ||||||   | ||||||||||| || ||       ||| |
Db      10450 ACACCGGATCGGGTGGATGTGGTGTTGCGGTCCAAGGTGGACGCGGCGTGGCACCTGCAC 10391

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||| |||| || |   ||||  |||||||||| ||  |||| || |||   || ||| |
Db      10390 GAGTTGACTCGCGACCTGGATGTGTCGGCGTTTGTCATGTTTTCGTCGATGGCCGGGCTG 10331

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | || ||| | || |||| |||||| || || |||||| |    | |||| ||||||||
Db      10330 GTCGGATCGTCGGGCCAGGCCAACTATGCGGCCGCCAATTCGTTTTTGGATGCGCTGGCC 10271

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||  ||||||||    || |||||||   | || ||||  ||||| ||||||
Db      10270 GCCCACCGGCGGGCCCATGGGCTGCCGGCCATCTCCCTGGGCTGGGGTCTGTGG 10217


Search completed: June 17, 2003, 17:55:51
Job time : 1038.31 secs
```

DR WPI; 2001-267102/28.
XX
PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives
PT -
XX
PS Claim 7; Page 92-102; 354pp; German.
XX
CC This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (II); (ii) to generate a library of polyketide synthases; (iii) for
CC adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC aglycone; and (iv) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC macrolides with insecticidal, but not antibacterial, activity, and can
CC also be used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (II) at significantly increased levels or produce new derivatives of
CC (II). This sequence represents a genomic DNA fragment of the S. spinosa
CC genome which contains the coding regions for proteins involved in
CC forosamine and trimethylrhamnose biosynthesis.
XX
SQ Sequence 29736 BP; 4401 A; 10346 C; 10080 G; 4909 T; 0 other;

Query Match 44.9%; Score 239.6; DB 22; Length 29736;
Best Local Similarity 65.5%; Pred. No. 8.1e-36;
Matches 350; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

```
Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  | ||| |||| || || || || || | ||| || ||   | | || || ||||||
Db       3604 GGCACCGTGCTGGTCACCGGCGGAACCGGCGGGCTCGGGGCCCTGCTCGCACGCCATCTG 3663

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |   | |      || |||||| | ||| || | |  || || ||||||| ||  || ||
Db       3664 GTCAGCGCCCACGGTGTGCGGCACCTGTTGCTCGCAAGTCGCCGTGGTCTGGCGGCGCCC 3723

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || |||||  |  |||  ||| |||||||    |   || || || ||   | | || |||
Db       3724 GGAGCGGATGAGCTGGTGGCCGAGCTGGAGCAGGCCGGCGCCGACGTCGCGGTCGTCGCG 3783

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              || ||   ||  || |||      | | |  |||||| ||| |   || |||||  |
Db       3784 TGCGACTCGGCAGATCGGGACTCGCTTGCGCGGCTGGTGGCGTCGGTGCCTGCGGAAAAC 3843

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||| ||  ||  || ||||| || || ||||||||||| || |    | |    ||  |
Db       3844 CCGTTGCGGGTGGTGGTGCACGCCGCCGGTGTGCTGGATGACGGTGTGCTGATGTCGATG 3903

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
               |||| |||||  ||| | |||||||   | |||| ||||||||||    |  ||| |
Db       3904 TCGCCGGAGCGCTTGGACGCGGTGTTGCGGCCCAAAGTGGATGCCGCGTGGTACCTGCAC 3963

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||||||| ||||   | |  |||||||||||||||| |||||||||||| ||||||  |
Db       3964 GAGCTGACTCGGGAACTCGGTCTGTCGGCGTTCGTGTTGTTCTCCTCGGTCGCGGGCCTG 4023

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | ||    || |||||| |||| ||||| || | ||| ||    ||||| ||  |||||
Db       4024 TTCGGCGGTGCGGGGCAGAGCAATTACGCTGCCGGCAACGCTTTCCTGGATGCCTTGGCG 4083

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              |  | |||| |||   ||| ||||| | | |||||||||| |  |||||||||
Db       4084 CATTGCCGGCAGGCCCAGGGGCTGCCCGCGCTGTCGCTGGCCTCCGGGCTGTGG 4137
```

Search completed: June 17, 2003, 13:49:04
Job time : 123.318 secs

```
Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              || || | |  || || |||||||||||||||||||| || |  | | | |   |  |
Db      10892 GCGGTGTCCGCGGTGGTTCATGCGGCTGGTGTGCTGGATGACGGCATGGTGGGTGCGTTG 10951

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||    |||||| ||  | ||||| |   | | ||||||||||| |         ||  ||
Db      10952 ACCGGGGAGCGGTTGTCCGCGGTGCTGCGGCCGAAGGTGGATGCTGTCTGGCATCTACAT 11011

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||  ||| || ||  |||| |||   ||||||||  | ||||||||   ||| ||| ||
Db      11012 GAGGCGACCCGCGGCCTGGACCTGGACGCGTTCGTCGTCTTCTCCTCCCTCGCCGGGGTC 11071

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | ||     |||| |||| |||||||||| || || || |||   ||||||||||||  |
Db      11072 TTCGGCAGTCCCGGCCAGGCCAACTACGCGGCCGCGAACGCCTTCCTGGACGCGCTGATG 11131

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                |||| |||||| ||| ||||| ||  |||| || || ||||| | ||||
Db      11132 ACGCGGCGCCGGGCGGAGGGACTGCCCGGCCTGTCACTCGCATGGGGACCGTGG 11185
```

RESULT 14
AAF88339
ID AAF88339 standard; DNA; 16767 BP.
XX
AC AAF88339;
XX
DT 28-AUG-2001 (first entry)
XX
DE S. spinosa DNA fragment encoding ORF22, SEQ ID 49.
XX
KW Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; polyketide synthase; ds.
XX
OS Saccharopolyspora spinosa.
XX
PN DE19957268-A1.
XX
PD 08-MAR-2001.
XX
PF 29-NOV-1999; 99DE-1057268.
XX
PR 27-AUG-1999; 99DE-1040596.
XX
PA (FARB) BAYER AG.
XX
PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
XX
DR WPI; 2001-267102/28.
DR P-PSDB; AAB70969.
XX
PT New nucleic acid encoding enzymes for spinosyn biosynthesis, useful for
PT recombinant production of insecticidal spinosyns and their derivatives
PT -
XX
PS Claim 7a; Page 284-313; 354pp; German.
XX
CC This invention describes a novel method nucleic acid (I) and its encoded
CC polypeptide (II) containing at least one region that encodes an enzymatic
CC activity involved in biosynthesis of spinosyns. (I) are used (i) to
CC identify, inactivate or modulate genes involved in the biosynthesis of
CC (II); (ii) to generate a library of polyketide synthases; (iii) for
CC adding forosamine or trimethylrhamnose to a spinosyn or polyketide
CC aglycone; and (iv) for recombinant production of the corresponding
CC enzymes, which are used for production of (II), their precursors or
CC derivatives, including production of transgenic plants that express (II)
CC and thus have increased resistance to insects. (I) are also useful as
CC markers for sequencing of the Saccharopolyspora spinosa genome. (II) are
CC macrolides with insecticidal, but not antibacterial, activity, and can
CC also be used to raise specific antibodies, useful for identifying
CC expression clones in a gene bank. Cells transformed with (I) may produce
CC (II) at significantly increased levels or produce new derivatives of
CC (II). This sequence, ORF 22, encodes an S. spinosa polyketide synthase.
XX
SQ Sequence 16767 BP; 2256 A; 4805 C; 6457 G; 3249 T; 0 other;

```
  Query Match             44.9%;  Score 239.6;  DB 22;  Length 16767;
  Best Local Similarity   65.5%;  Pred. No. 8.2e-36;
  Matches  350;  Conservative    0;  Mismatches  184;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  | ||| |||| || || || || || | ||| || ||   | | || || ||||||
Db      14797 GGCACCGTGCTGGTCACCGGCGGAACCGGCGGGCTCGGGGCCCTGCTCGCACGCCATCTG 14856

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |   | |      || ||||||| | ||| || | |  || || ||||||| ||  || ||
Db      14857 GTCAGCGCCCACGGTGTGCGGCACCTGTTGCTCGCAAGTCGCCGTGGTCTGGCGGCGCCC 14916

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || ||||| |  |||  ||| |||||||     |   || || || ||   | | || |||
Db      14917 GGAGCGGATGAGCTGGTGGCCGAGCTGGAGCAGGCCGGCGCCGACGTCGCGGTCGTCGCG 14976

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              || ||   ||  || |||     | | |  |||||| ||| |   || ||||| |
Db      14977 TGCGACTCGGCAGATCGGGACTCGCTTGCGCGGCTGGTGGCGTCGGTGCCTGCGGAAAAC 15036

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||| ||  ||  || ||||| || || |||||||||||| || |    | |    ||  |
Db      15037 CCGTTGCGGGTGGTGGTGCACGCCGCCGGTGTGCTGGATGACGGTGTGCTGATGTCGATG 15096

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |||| |||||  ||| | |||||||   | |||| ||||||||||     |  ||| |
Db      15097 TCGCCGGAGCGCTTGGACGCGGTGTTGCGGCCCAAAGTGGATGCCGCGTGGTACCTGCAC 15156

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||||||| ||||   | |  |||||||||||||| |||||||||||| ||||||  |
Db      15157 GAGCTGACTCGGGAACTCGGTCTGTCGGCGTTCGTGTTGTTCTCCTCGGTCGCGGGCCTG 15216

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | ||     || |||||| |||| |||||.|| | ||| ||     ||||| ||  |||||
Db      15217 TTCGGCGGTGCGGGGCAGAGCAATTACGCTGCCGGCAACGCTTTCCTGGATGCCTTGGCG 15276

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              |  | |||| |||   ||| ||||| | | |||||||||| |  |||||||||
Db      15277 CATTGCCGGCAGGCCCAGGGGCTGCCCGCGCTGTCGCTGGCCTCCGGGCTGTGG 15330
```

RESULT 15
AAF88317
ID AAF88317 standard; DNA; 29736 BP.
XX
AC AAF88317;
XX
DT 28-AUG-2001 (first entry)
XX
DE S. spinosa DNA fragment SEQ ID 6.
XX
KW Forosamine; trimethylrhamnose; polyketide synthase; biosynthesis;
KW spinosyn; polyketide aglycone; transgenic plant; insect resistance;
KW macrolide; insecticidal; ds.
XX
OS Saccharopolyspora spinosa.
XX
PN DE19957268-A1.
XX
PD 08-MAR-2001.
XX
PF 29-NOV-1999; 99DE-1057268.
XX
PR 27-AUG-1999; 99DE-1040596.
XX
PA (FARB) BAYER AG.
XX
PI Eberz G, Moehrle V, Froede R, Velten R, Salas JA;
XX

```
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

  Query Match             46.7%;  Score 249.2;  DB 22;  Length 4403765;
  Best Local Similarity   66.7%;  Pred. No. 1.2e-37;
  Matches  356;  Conservative    0;  Mismatches  178;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||||||| | || |||||| || ||  ||  |||| ||||||||||| || ||  ||
Db    2298157 GGCACGGTGTTGATCACCGGTGGCACCGGGATGGCGGGTTCGGCGGTGGCCCGTCACGTG 2298098

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  | |      |||||||||  ||||| ||||||||||||||||| || ||||||||||
Db    2298097 GTGGCTCGTCATGGGGTGCGCAATCTGGTGTTGGTGAGCCGGCGCGGCCCGGATGCTCCC 2298038

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || || | ||  |||  ||| ||| |||||||     || ||| |||| | | | || ||
Db    2298037 GGGGCTGCGGAGCTGGTGGCCGAGTTGGCCGCCGCCGGTGCGCAGGTACAGGTGGTGGCC 2297978

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||| | || ||| ||   || | ||| |  | || | |  | | |||| |    |  |
Db    2297977 TGTGACGCGGCGGATCGAGCGGCGTTGGCCAAGGTGATCGCCGATATTCCGGTTCAGCAT 2297918

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||  || |||| ||  | ||  | || || |  || ||||| |||  | |    || ||
Db    2297917 CCATTGTCGGGCGTGATCCACACCGCCGGCGCACTCGACGACGCGGTGGTGATGTCACTG 2297858

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              || || || ||| |||    ||||||   | ||||||||| || ||       ||| |
Db    2297857 ACACCGGATCGGGTGGATGTGGTGTTGCGGTCCAAGGTGGACGCGGCGTGGCACCTGCAC 2297798

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||| |||| || |   ||||  |||||||||| ||  |||| || |||   || ||| |
Db    2297797 GAGTTGACTCGCGACCTGGATGTGTCGGCGTTTGTCATGTTTTCGTCGATGGCCGGGCTG 2297738

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | || ||| | || |||| |||||| || || |||||| |    | |||| ||||||||
Db    2297737 GTCGGATCGTCGGGCCAGGCCAACTATGCGGCCGCCAATTCGTTTTTGGATGCGCTGGCC 2297678

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                ||  |||||||||    || |||||||    | || ||||  |||||  ||||||
Db    2297677 GCCCACCGGCGGGCCCATGGGCTGCCGGCCATCTCCCTGGGCTGGGGTCTGTGG 2297624


RESULT 13
AAD17185
ID   AAD17185 standard; DNA; 27541 BP.
XX
AC   AAD17185;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys2; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (454..1191)
FT                   /*tag=  a
FT                   /product= "NysF protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (1275..3092)
FT                   /*tag=  b
FT                   /product= "NysG protein"
FT   CDS             complement (3070..4824)
FT                   /*tag=  c
FT                   /note= "CDS does not include start codon"
FT                   /product= "NysH protein"
FT   CDS             5122..6156
FT                   /*tag=  d
FT                   /product= "NysD3 protein"
FT   CDS             6338..27541
FT                   /*tag=  e
FT                   /product= "NysI partial protein"
FT                   /note= "CDS does not include stop codon"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 151-166; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

  Query Match             45.7%;  Score 244;  DB 22;  Length 27541;
  Best Local Similarity   67.8%;  Pred. No. 1.2e-36;
  Matches  362;  Conservative    0;  Mismatches  160;  Indels   12;  Gaps    1;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||  |||| || ||||| |||||||  |||||||||| | |||||||  ||||||
Db     10664 GGTGCGGTCCTGGTCACTGGTGGTACGGGTGGTCTGGGTGCGGTGTTGGCGCGTCATCTG 10723

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  |  | || ||||||||||||||||||||||||| || ||  |||||      |||   |
Db     10724 GTGGCCGAGTATGGGGTGCGGGATCTGCTGTTGGTCAGTCGCAGTGGTGAACGTGCCGTG 10783

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              ||||| | ||   ||  ||||||||| || |   |||| |||  ||||||| | ||||||
Db     10784 GGTGCTGGGGAGTTGGTGGCGGAGCTTGCGGGTGTGGGTGCGCGGGTGCGGGTGGTTGCG 10843

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              |||||||||   || ||    | ||||||   | || |            || |||| |
Db     10844 TGTGATGTGACCGATCGTGCCGCGGTGGTGGAGTTGGT-----------TGGCGGGCAT 10891
```

```
Db      13224 GAACTGACCCGGACCAGCCCGCTGTCGGTGTTCGCGGTCTTCTCGGGCGCTGCCGGCATC 13283

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480

Db      13284 CTGGGCCGCCCCCGGACAGGCCAACTACGCCGCCGCCAACACCTTCCTCGACGCGCTCGCG 13343

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534

Db      13344 CAGCACCGCCGCGCCCACGGCCTCCCCGCCGTGTCGCTGGCCTGGGGGCTGTGG 13397


RESULT 11
AAI99682/c
ID   AAI99682 standard; DNA; 4411529 BP.
XX
AC   AAI99682;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fleischmann RD,  White OR,  Fraser CM,  Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ  -
XX
PS   Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
CC   H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC   M. tuberculosis and has valuable application in the fields of
CC   tuberculosis genetics, epidemiology, patient treatment and epidemic
CC   monitoring.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html?DocID=6294328B1.
XX
SQ   Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

  Query Match            47.0%;  Score 250.8;  DB 22;  Length 4411529;
  Best Local Similarity   66.9%;  Pred. No. 6.1e-38;
  Matches  357;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60

Db    2295857 GGCACGGTGTTGATCACCGGTGGCACCGGGATGGCGGGTTCGGCGGTGGCCCGTCACGTG 2295798

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120

Db    2295797 GTGGCTCGTCATGGGGTGCGCAATCTGGTGTTGGTGAGCCGGCGCGGCCCGGATGCTCCC 2295738

Qy         121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180

Db    2295737 GGGGCTGCGGAGCTGGTGGCCGAGTTGGCCGCCGCCGGTGCGCAGGTGCAGGTGGTGGCC 2295678

Qy         181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240

Db    2295677 TGTGACGCGGCGGATCGAGCGGCGTTGGCCAAGGTGATCGCCGATATTCCGGTTCAGCAT 2295618

Qy         241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300

Db    2295617 CCATTGTCGGGCGTGATCCACACCGCCGGCGCACTCGACGACGCGGTGGTGATGTCACTG 2295558

Qy         301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360

Db    2295557 ACACCGGATCGGGTGGATGTGGTGTTGCGGTCCAAGGTGGACGCGGCGTGGCACCTGCAC 2295498

Qy         361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420

Db    2295497 GAGTTGACTCGCGACCTGGATGTGTCGGCGTTTGTCATGTTTTCGTCGATGGCCGGGCTG 2295438

Qy         421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480

Db    2295437 GTCGGATCGTCGGGCCAGGCCAACTATGCGGCCGCCAATTCGTTTTTGGATGCGCTGGCC 2295378

Qy         481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534

Db    2295377 GCCCACCGGCGGGCCCATGGGCTGCCGGCCATCTCCCTGGGCTGGGGTCTGTGG 2295324


RESULT 12
AAI99683/c
ID   AAI99683 standard; DNA; 4403765 BP.
XX
AC   AAI99683;
XX
DT   15-JAN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
XX
KW   Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
KW   variation; epidemiology; patient treatment; epidemic monitoring; ds.
XX
OS   Mycobacterium tuberculosis.
XX
PN   US6294328-B1.
XX
PD   25-SEP-2001.
XX
PF   24-JUN-1998;   98US-0103840.
XX
PR   24-JUN-1998;   98US-0103840.
XX
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fleischmann RD,  White OR,  Fraser CM,  Venter JC;
XX
DR   WPI; 2001-647261/74.
XX
PT   Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT   determining the nucleotide sequence of the strain at positions in the
PT   genome corresponding to positions where M. tuberculosis strains CDC
PT   1551 and H37Rv differ  -
XX
PS   Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
XX
CC   The invention relates to evaluating strain variation within and between
CC   different populations of the tuberculosis bacterial pathogen,
CC   Mycobacterium tuberculosis or related Mycobacterium by determining the
CC   nucleotide sequence of the first strain at positions in the complete
CC   sequence of the genome that correspond to positions that differ in the
CC   nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
```

```
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ   Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

  Query Match             48.2%;  Score 257.2;  DB 22;  Length 125401;
  Best Local Similarity   67.6%;  Pred. No. 4.4e-39;
  Matches  361;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||||||||||||||||| |||||||  |||||||  |  ||||||  |||||||||
Db      86180 GGAACGGTGTTGGTGACGGGTGGTACGGGTGGTCTGGGTGGGTTGGTGGCCCGGCATCTG 86239

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |   |   |||||||||||||| || || |||||    || ||   ||||  || |||| ||
Db      86240 GTGCGGTCGTGTGGGGTGCGGCATTTGTTGTTGACCAGTCGTTCTGGTGTGGGTGCTGCG 86299

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              |||||||  ||  ||   ||||||  |||   |||||||  |||  |||   | |||  |||
Db      86300 GGTGCGGCCGGGTTGGTCGCGGAGTTGGAGTCGTTGGGCGCGCGGGTTGTGGTTGCGGCG 86359

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||| ||  |   ||  || |     | || | |  |||||  | |     || |
Db      86360 TGTGATGTGGGTGATGGCTCGGCTGTTGCGGAGTTGGTTGCCGGTGTGTCGGAGTCGTAT 86419

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||| || | |  || ||||||||||||||||||| |||| || |     | | |  ||  |
Db      86420 CCGTTGTCTGCGGTGGTGCATGCGGCTGGTGTGTTGGATGACGGTGTGGTGGGTTCGTTG 86479

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||||| ||||||  |||   |||||||     | ||||||||||  ||       ||| ||
Db      86480 ACGCCGGAGCGGTTGGCTGCGGTGTTGCGTCCGAAGGTGGATGGTGCGTGGAACCTGCAT 86539

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||  |||||| ||| |||| |||   |||||  ||   | |||||  || |  |||||   |
Db      86540 GAGGCGACGCGTGGTCTGGATCTGGACGCGTTTGTTGTCTTCTCGTCTGTTGCGGGTGTG 86599

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | |||   ||  ||  |||| |||||| ||  ||||  |||||   | ||||||||  ||  |
Db      86600 TTCGGGGGTGCGGGTCAGGCCAACTATGCGGCGGGTAATGCGTTTTTGGACGCGTTGATG 86659

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                |  |||  ||| |  ||  |||| || ||||||  ||||||||||    |||
Db      86660 GTTCATCGGGTGGCTGGTGGGTTGCCTGGTGTGTCGTTGGCGTGGGGTGCTTGG 86713


RESULT 10
AAT80413
ID   AAT80413 standard; DNA; 43280 BP.
XX
AC   AAT80413;
XX
DT   27-FEB-1998  (first entry)
XX
DE   Tylactone synthase gene cluster.
XX
KW   Tylactone synthase gene cluster; tylG gene; multifunctional protein;
KW   polyketide; tylactone synthesis; antibiotic; tylosin; ss.
XX
OS   Streptomyces fradiae.
XX
FH   Key             Location/Qualifiers
FT   CDS             816..14243
FT                   /*tag=  a
FT                   /transl_except= (pos: 816..818, aa: Met)
FT                   /note= "ORF1 encodes protein shown in AAW22601"
FT   CDS             14351..19945
FT                   /*tag=  b
FT                   /transl_except= (pos: 14351..14353, aa: Met)
FT                   /note= "ORF2 encodes protein shown in AAW22602"
FT   CDS             20010..31199
FT                   /*tag=  c
FT                   /transl_except= (pos: 20010..20012, aa: Met)
FT                   /note= "ORF3 encodes protein shown in AAW22603"
FT   CDS             31232..36067
FT                   /*tag=  d
FT                   /note= "ORF4 encodes protein shown in AAW22604"
FT   CDS             36249..41774
FT                   /*tag=  e
FT                   /note= "ORF5 encodes protein shown in AAW22605"
XX
PN   EP791655-A2.
XX
PD   27-AUG-1997.
XX
PF   19-FEB-1997;   97EP-0301056.
XX
PR   22-FEB-1996;   96US-0012078.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Dehoff BS,  Kuhstoss SA,  Rosteck PR,  Sutton KL;
XX
DR   WPI; 1997-418046/39.
DR   P-PSDB; AAW22601-W22605.
XX
PT   DNA encoding Streptomyces fradiae tylactone synthase domain - for
PT   production of tylosin-related polyketide compounds
XX
PS   Claim 2; Pages 8-66; 220pp; English.
XX
CC   This sequence represents the tylactone synthase gene cluster of the
CC   invention. This sequence is also referred to as the tylG gene, and was
CC   isolated from Streptomyces fradiae. This sequence encodes multifunctional
CC   proteins which direct the synthesis of the polyketide tylactone, isolated
CC   from Streptomyces fradiae. Tylactone is the basic building block of the
CC   antibiotic tylosin. The DNA sequence can be modified so as to alter the
CC   type of carboxylic acids incorporated, the number of carboxylic acids
CC   incorporated and/or the post-condensation reactions performed, thereby
CC   resulting in novel tylosin-related polyketides.
XX
SQ   Sequence 43280 BP; 5269 A; 15187 C; 16651 G; 6173 T; 0 other;

  Query Match             47.0%;  Score 250.8;  DB 18;  Length 43280;
  Best Local Similarity   66.9%;  Pred. No. 6.9e-38;
  Matches  357;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||  |||||||||| | ||| ||   ||| ||    |  | ||||  |||||||||
Db      12864 GGCACGGTCCTGGTGACGGGCGCGACCGGCACGCTCGGGCGGCTGCTGGCCCGGCATCTG 12923

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              || |  | |   || |||||  |||||||| || |||||||  |  || | || |||
Db      12924 GCCGCCGAGCACGGTGTGCGCCATCTGCTGCTGCTGAGCCGCGGCGGCCGGGCTGCCGAA 12983

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              || ||||| |  ||    ||||| |||||||  |||||  ||  |||  | |  | |  ||
Db      12984 GGCGCGGACGAACTCGCCGCGGAACTGGCCGGGTTGGAAGCCGAGCCGTGCTTCGCGGCC 13043

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||| | || ||| ||   ||   |||  ||| |||||| ||   || || ||||   |
Db      13044 TGTGACGCGGCGGACCGCGAGGCCCTGGCACGGGTGCTGGCGGAGGTGCCGGCCGACCGG 13103

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||  || ||  | || ||||| || ||||| |||||  |   ||  ||| | | ||
Db      13104 CCGCTGACCGGAGTGATCCACGCGGCCGGGGTGCTCGACGACGGCACACTCGACGCGCTG 13163

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              || || || ||| | || || ||  |   | | |||| ||| || || ||     ||| |
Db      13164 ACCCCGGAACGGATCGGTACCGTCATGCGGCCGAAGGCGGACGCGGCGCTGAACCTGCAC 13223

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              || ||||| |||   |    |||||||| ||||| | | |||||    || || || |||
```

```
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10125, AAE10126, AAE10127, AAE10128, AAE10129, AAE10130,
DR   AAE10131, AAE10132, AAE10133, AAE10134, AAE10135, AAE10136, AAE10137.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 2; Page 116-151; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
SQ   Sequence 65140 BP; 8270 A; 25171 C; 22273 G; 9426 T; 0 other;

  Query Match             48.2%;  Score 257.2;  DB 22;  Length 65140;
  Best Local Similarity   67.6%;  Pred. No. 4.5e-39;
  Matches  361;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||||||||||||||||| |||||||  ||||||| |  |||||| |||||||||
Db      25919 GGAACGGTGTTGGTGACGGGTGGTACGGGTGGTCTGGGTGGGTTGGTGGCCCGGCATCTG 25978

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |   |  |||||||||||||| || || |||||   || ||   ||||  || |||| ||
Db      25979 GTGCGGTCGTGTGGGGTGCGGCATTTGTTGTTGACCAGTCGTTCTGGTGTGGGTGCTGCG 26038

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              |||||||  ||  ||   |||||| |||   ||||||| |||  |||   | |||  |||
Db      26039 GGTGCGGCCGGGTTGGTCGCGGAGTTGGAGTCGTTGGGCGCGCGGGTTGTGGTTGCGGCG 26098

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||| ||  |   ||  || |     | || | |  |||||  | |     || |
Db      26099 TGTGATGTGGGTGATGGCTCGGCTGTTGCGGAGTTGGTTGCCGGTGTGTCGGAGTCGTAT 26158

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||| || | |  || ||||||||||||||||||| |||| ||  |     | | |  ||  |
Db      26159 CCGTTGTCTGCGGTGGTGCATGCGGCTGGTGTGTTGGATGACGGTGTGGTGGGTTCGTTG 26218

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||||| |||||| |||   |||||||     | ||||||||||  ||       ||| ||
Db      26219 ACGCCGGAGCGGTTGGCTGCGGTGTTGCGTCCGAAGGTGGATGGTGCGTGGAACCTGCAT 26278

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||  ||||||  ||| |||| |||    |||||  ||   | |||||  || |  |||||  |
Db      26279 GAGGCGACGCGTGGTCTGGATCTGGACGCGTTTGTTGTCTTCTCGTCTGTTGCGGGTGTG 26338

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
               | |||    ||  || |||| |||||| || ||||  |||||    | |||||||| || |
Db      26339 TTCGGGGGTGCGGGTCAGGCCAACTATGCGGCGGGTAATGCGTTTTTGGACGCGTTGATG 26398

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                 |  |||  ||| |  ||  |||| || |||||| ||||||||||    |||
Db      26399 GTTCATCGGGTGGCTGGTGGGTTGCCTGGTGTGTCGTTGGCGTGGGGTGCTTGG 26452


RESULT 9
AAD17186
ID   AAD17186 standard; DNA; 125401 BP.
XX
AC   AAD17186;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             6337..34771
FT                   /*tag=  a
FT                   /product= "NysI complete protein"
FT   CDS             34792..51099
FT                   /*tag=  b
FT                   /product= "NysJ protein"
FT   CDS             51155..57355
FT                   /*tag=  c
FT                   /product= "NysK protein"
FT   CDS             57503..58687
FT                   /*tag=  d
FT                   /product= "NysL protein"
FT   CDS             complement (58786..58980)
FT                   /*tag=  e
FT                   /product= "NysM protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (59045..60241)
FT                   /*tag=  f
FT                   /product= "NysN protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (60238..61296)
FT                   /*tag=  g
FT                   /product= "NysD2 complete protein"
FT   CDS             120628..121308
FT                   /*tag=  h
FT                   /product= "NysR4 (long) protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR   AAE10149, AAE10150.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Claim 1; Page 188-254; 266pp; English.
XX
```

```
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
PA   (ALPH-) ALPHARMA AS.
PA   (SINV-) SINVENT AS.
PA   (DZIE/) DZIEGLEWSKA H.
PA   (ZOTC/) ZOTCHEV S B.
PA   (SEKU/) SEKUROVA O N.
PA   (FJAE/) FJAERVIK E.
PA   (BRAU/) BRAUTASET T.
PA   (STRO/) STROM A R.
XX
PI   Zotchev SB,  Sekurova ON,  Fjaervik E,  Brautaset T,  Strom AR;
PI   Valla S,  Ellingsen TE,  Sletta H,  Gulliksen O;
XX
DR   WPI; 2001-557614/62.
DR   P-PSDB; AAE10153.
XX
PT   New nystatin polyketide synthase polynucleotides and polypeptides,
PT   useful as antibiotics and antifungals -
XX
PS   Example 2; Page 186-187; 266pp; English.
XX
CC   The present invention relates to the cloning and sequencing of the gene
CC   cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC   involved in the biosynthesis of the macrolide antibiotic nystatin.
CC   The nystatin PKS is useful as antifungal antibiotics. The present
CC   sequence is a ERD48 insert DNA in pKR48 which is used in an
CC   exemplification of the invention.
XX
SQ   Sequence 2700 BP; 324 A; 1087 C; 876 G; 413 T; 0 other;

  Query Match             48.2%;  Score 257.2;  DB 22;  Length 2700;
  Best Local Similarity   67.6%;  Pred. No. 4.9e-39;
  Matches  361;  Conservative    0;  Mismatches  173;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||  ||||||||||||||||||| |||||||  |||||||  |  ||||||  |||||||||
Db       1170 GGAACGGTGTTGGTGACGGGTGGTACGGGTGGTCTGGGTGGGTTGGTGGCCCGGCATCTG 1229

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              |  |   |||||||||||||| || || |||||    || ||    ||||   || |||| ||
Db       1230 GTGCGGTCGTGTGGGGTGCGGCATTTGTTGTTGACCAGTCGTTCTGGTGTGGGTGCTGCG 1289

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              |||||||  ||  ||   |||||| |||   ||||||| |||  |||   | |||  |||
Db       1290 GGTGCGGCCGGGTTGGTCGCGGAGTTGGAGTCGTTGGGCGCGCGGGTTGTGGTTGCGGCG 1349

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||| ||  |   ||  || |     | || | |  |||||  | |    || |
Db       1350 TGTGATGTGGGTGATGGCTCGGCTGTTGCGGAGTTGGTTGCCGGTGTGTCGGAGTCGTAT 1409

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||| || | |  || ||||||||||||||||||| |||| ||  |     | | |  ||  |
Db       1410 CCGTTGTCTGCGGTGGTGCATGCGGCTGGTGTGTTGGATGACGGTGTGGTGGGTTCGTTG 1469

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              |||||  |||||| |||   |||||||     | ||||||||||  ||        ||| ||
Db       1470 ACGCCGGAGCGGTTGGCTGCGGTGTTGCGTCCGAAGGTGGATGGTGCGTGGAACCTGCAT 1529

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              |||  |||||| ||| |||| |||   |||||  ||   | |||||  || |  |||||  |
Db       1530 GAGGCGACGCGTGGTCTGGATCTGGACGCGTTTGTTGTCTTCTCGTCTGTTGCGGGTGTG 1589

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              | |||   || || |||| |||||| || ||||  |||||   | ||||||||  ||  |
Db       1590 TTCGGGGGTGCGGGTCAGGCCAACTATGCGGCGGGTAATGCGTTTTTGGACGCGTTGATG 1649

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                |  |||  ||| |  ||  |||| || |||||| ||||||||||    |||
Db       1650 GTTCATCGGGTGGCTGGTGGGTTGCCTGGTGTGTCGTTGGCGTGGGGTGCTTGG 1703
```

```
RESULT 8
AAD17184
ID   AAD17184 standard; DNA; 65140 BP.
XX
AC   AAD17184;
XX
DT   29-NOV-2001  (first entry)
XX
DE   Streptomyces noursei nys1 DNA of nystatin PKS gene cluster.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; nys1; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             complement (1..1035)
FT                   /*tag= a
FT                   /product= "NysD2 partial protein"
FT                   /note= "CDS does not include stop codon"
FT   CDS             complement (1056..2576)
FT                   /*tag= b
FT                   /product= "NysD1 protein"
FT   CDS             2806..6906
FT                   /*tag= c
FT                   /product= "NysA protein"
FT   CDS             6952..16530
FT                   /*tag= d
FT                   /product= "NysB protein"
FT   CDS             16550..49840
FT                   /*tag= e
FT                   /product= "NysC protein"
FT   CDS             50260..51015
FT                   /*tag= f
FT                   /product= "NysE protein"
FT   CDS             51405..54305
FT                   /*tag= g
FT                   /product= "NysR1 protein"
FT   CDS             54329..57190
FT                   /*tag= h
FT                   /product= "NysR2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             57180..59963
FT                   /*tag= i
FT                   /product= "NysR3 protein"
FT   CDS             60415..61047
FT                   /*tag= j
FT                   /product= "NysR4 (short) protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             61736..62497
FT                   /*tag= k
FT                   /product= "NysR5 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             complement (62551..63615)
FT                   /*tag= l
FT                   /product= "ORF2 protein"
FT                   /note= "CDS does not include start codon"
FT   CDS             63765..64961
FT                   /*tag= m
FT                   /product= "ORF1 protein"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
XX
PA   (UYNO-) UNIV NORGES TEKNISK NATURVITENSKAPELIGE.
PA   (SNTF ) SINTEF STIFTELSEN IND TEK FORSK.
```

```
Db      19525 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 19584

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19585 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 19644

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19645 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 19704

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19705 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 19764

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19765 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 19818


RESULT 6
AAH79278
ID   AAH79278 standard; DNA; 31422 BP.
XX
AC   AAH79278;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 2.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag=  a
FT                   /product= "AAG65266"
FT   CDS             14824..31422
FT                   /*tag=  b
FT                   /product= "AAG65267"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65266, AAG65267.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Disclosure; Page 103-149; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 31422 BP; 4136 A; 10237 C; 11677 G; 5372 T; 0 other;

  Query Match             99.7%;  Score 532.4;  DB 22;  Length 31422;
  Best Local Similarity   99.8%;  Pred. No. 8.3e-90;
  Matches  533;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19285 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 19344

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19345 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 19404

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db      19405 GGTGCGGAGGGTTTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 19464

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19465 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 19524

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19525 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 19584

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19585 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 19644

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19645 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 19704

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19705 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 19764

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19765 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 19818


RESULT 7
AAD17193
ID   AAD17193 standard; DNA; 2700 BP.
XX
AC   AAD17193;
XX
DT   29-NOV-2001  (first entry)
XX
DE   ERD48 insert DNA in pKR48.
XX
KW   Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
KW   antifungal; antibiotic; ds.
XX
OS   Streptomyces noursei.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..2699
FT                   /*tag=  a
FT                   /product= "Protein encoded by ERD48 insert DNA"
XX
PN   WO200159126-A2.
XX
PD   16-AUG-2001.
XX
PF   08-FEB-2001; 2001WO-GB00509.
XX
PR   08-FEB-2000; 2000GB-0002840.
PR   10-APR-2000; 2000GB-0008786.
PR   14-APR-2000; 2000GB-0009387.
```

```
XX
PI   Endo H,  Yamaguchi H,  Kanda Y,  Hashimoto S,  Omura S,  Ikeda H;
XX
DR   WPI; 2001-582053/65.
DR   P-PSDB; AAG65264, AAG65265.
XX
PT   New modified avermectin aglycone synthase derived from Streptomyces
PT   avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT   drugs and pesticides -
XX
PS   Example 2; Page 58-123; 257pp; Japanese.
XX
CC   The present invention relates to the production of modified derivatives
CC   of avermectin aglycone synthase (AAS) derived from Streptomyces
CC   avermitilis. The activity of an acyl carrier protein (ACP),
CC   beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC   beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC   reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC   suppressed. The process can be used in the production of drugs, vetinary
CC   drugs and pesticides. The present sequence is a fragment of the S.
CC   avermitilis genome.
XX
SQ   Sequence 30690 BP; 5355 A; 12455 C; 8617 G; 4263 T; 0 other;

  Query Match            100.0%;  Score 534;  DB 22;  Length 30690;
  Best Local Similarity  100.0%;  Pred. No. 4.2e-90;
  Matches  534;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10609 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 10668

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10669 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 10728

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10729 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 10788

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10789 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 10848

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10849 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 10908

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10909 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 10968

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10969 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 11028

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11029 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 11088

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11089 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 11142


RESULT 5
AAA92302
ID   AAA92302 standard; DNA; 31422 BP.
XX
AC   AAA92302;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAII SEQ ID NO:2.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..14646
FT                   /*tag= a
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             14824..31422
FT                   /*tag= b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23751, AAB23752.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 2; Page 134-203; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 31422 BP; 4136 A; 10238 C; 11677 G; 5371 T; 0 other;

  Query Match            99.7%;  Score 532.4;  DB 21;  Length 31422;
  Best Local Similarity  99.8%;  Pred. No. 8.3e-90;
  Matches  533;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19285 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 19344

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19345 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 19404

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      19405 GGTGCGGAGGGTTTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 19464

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      19465 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 19524

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11786 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 11845

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11846 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 11905

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11906 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 11959


RESULT 3
AAA92301
ID   AAA92301 standard; DNA; 30690 BP.
XX
AC   AAA92301;
XX
DT   10-JAN-2001  (first entry)
XX
DE   S. avermitilis avermectin aglycon synthase DNA aveAI SEQ ID NO:1.
XX
KW   Streptomyces avermitilis; avermectin aglycon synthase; biosynthesis;
KW   multifunctional enzyme; polyketide; avermectin; veterinary drug;
KW   agrochemical; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag=  a
FT                   /note= "avermectin aglycon synthase protein"
FT   CDS             11971..30690
FT                   /*tag=  b
FT                   /note= "avermectin aglycon synthase protein"
XX
PN   WO200050605-A1.
XX
PD   31-AUG-2000.
XX
PF   23-FEB-2000; 2000WO-JP01041.
XX
PR   24-FEB-1999;   99JP-0046961.
XX
PA   (KITA ) KITASATO INST.
XX
PI   Omura S,  Ikeda H;
XX
DR   WPI; 2000-565458/52.
DR   P-PSDB; AAB23749, AAB23750.
XX
PT   Avermectin aglycone synthase DNA and proteins encoded by all or part of
PT   it for the production of avermectin and its derivatives for drug and
PT   agrochemical use  -
XX
PS   Claim 2; Page 66-134; 314pp; Japanese.
XX
CC   The present sequence represents DNA which encodes avermectin aglycon
CC   synthase proteins. Also described are: (1) polypeptides encoded by all
CC   or part of the DNA; (2) expression vectors containing the DNA; (3) host
CC   cells transformed by the vectors; (4) preparation of the polypeptides
CC   by culture of the transformants; (5) preparation of avermectin aglycon
CC   or its derivatives by culture of transformed avermectin-producing
CC   microorganisms; and (6) oligonucleotides of 5-60 bases in length
CC   containing sense or antisense sequences from the avermectin aglycon
CC   synthase DNA. The enzymes are useful for the production of modified
CC   forms of avermectin and of the intermediates in its biosynthesis, for
CC   use as drugs, veterinary drugs and agrochemicals.
XX
SQ   Sequence 30690 BP; 5356 A; 12454 C; 8617 G; 4263 T; 0 other;

  Query Match           100.0%;  Score 534;  DB 21;  Length 30690;
  Best Local Similarity  100.0%;  Pred. No. 4.2e-90;
  Matches  534;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10609 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 10668

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10669 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 10728

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10729 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 10788

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10789 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 10848

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10849 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 10908

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10909 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 10968

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10969 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 11028

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11029 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 11088

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11089 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 11142


RESULT 4
AAH79277
ID   AAH79277 standard; DNA; 30690 BP.
XX
AC   AAH79277;
XX
DT   04-DEC-2001  (first entry)
XX
DE   Streptomyces avermitilis coding sequences SEQ ID NO: 1.
XX
KW   Avermectin aglycone synthase; AAS; avermectin derivative;
KW   drug production; vetinary drug; pesticide; ds.
XX
OS   Streptomyces avermitilis.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..11919
FT                   /*tag=  a
FT                   /product= "AAG65264"
FT                   /partial
FT   CDS             11971..30690
FT                   /*tag=  b
FT                   /product= "AAG65265"
XX
PN   WO200162939-A1.
XX
PD   30-AUG-2001.
XX
PF   23-FEB-2001; 2001WO-JP01381.
XX
PR   24-FEB-2000; 2000JP-0047405.
XX
PA   (KYOW ) KYOWA HAKKO KOGYO KK.
PA   (KITA ) KITASATO INST.
```

PI Endo H, Yamaguchi H, Kanda Y, Hashimoto S, Omura S, Ikeda H;
XX
DR WPI; 2001-582053/65.
DR P-PSDB; AAG65268.
XX
PT New modified avermectin aglycone synthase derived from Streptomyces
PT avermectilis used in production of 22,23-dihydroavermectin B1a used in
PT drugs and pesticides -
XX
PS Claim 10; Page 149-167; 257pp; Japanese.
XX
CC The present invention relates to the production of modified derivatives
CC of avermectin aglycone synthase (AAS) derived from Streptomyces
CC avermitilis. The activity of an acyl carrier protein (ACP),
CC beta-ketoacyl carrier protein synthase (KS), acyltransferase (AT),
CC beta-ketoacyl carrier protein reductase (KR), dehydratase (DH), enoyl
CC reductase (ER) and/or thioesterase (TE) domain may be reduced or
CC suppressed. The process can be used in the production of drugs, vetinary
CC drugs and pesticides. The present sequence is modified version of a
CC fragment of the S. avermitilis genome.
XX
SQ Sequence 11916 BP; 1831 A; 4428 C; 3820 G; 1837 T; 0 other;

Query Match 100.0%; Score 534; DB 22; Length 11916;
Best Local Similarity 100.0%; Pred. No. 4.3e-90;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10609 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 10668

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10669 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 10728

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10729 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 10788

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10789 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 10848

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10849 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 10908

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10909 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 10968

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10969 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 11028

Qy        421 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11029 CTGGGGTCGGCCGGGCAGGGCAACTACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCG 11088

Qy        481 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 534
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11089 TACCGGCGGCGGGCGGCGGGTCTGCCGGGGGTGTCGCTGGCGTGGGGGCTGTGG 11142
```

RESULT 2
AAZ58381
ID AAZ58381 standard; DNA; 12381 BP.
XX
AC AAZ58381;
XX
DT 23-MAY-2000 (first entry)
XX
DE Streptomyces avermitilis avermectin polyketide synthase modules 1+2.
XX
KW Polyketide synthase; avermectin; insecticide; ss.
XX
OS Streptomyces avermitilis.
XX
PN WO200001827-A2.
XX
PD 13-JAN-2000.
XX
PF 06-JUL-1999; 99WO-GB02158.
XX
PR 06-JUL-1998; 98GB-0014622.
XX
PA (BIOT-) BIOTICA TECHNOLOGY LTD.
PA (PFIZ) PFIZER INC.
XX
PI Kellenberger JL, Leadlay PF, Staunton J, Stutzman-Engwall KJ;
PI McArthur HAI;
XX
DR WPI; 2000-182117/16.
XX
PT Mutated Type I polyketide synthase containing a polylinker site in an
PT extension module for replacement of a reductive loop sequence, for
PT producing polyketides, e.g. B1 avermectin -
XX
PS Disclosure; Fig 7a-f; 75pp; English.
XX
CC The present sequence is that of DNA encoding the first 2 modules
CC of the avermectin polyketide synthase (PKS) of Streptomyces
CC avermitilis. The invention relates to nucleic acids encoding a
CC Type I PKS such as avermectin in which a polylinker with multiple
CC restriction sites replaces or 1 more PKS genes encoding enzymes
CC associated with reduction. Novel PKS are provided in which in
CC which the reductive loop in a selected module of the Type I PKS is
CC replaced with the equivalent segment from the same or different
CC PKS gene cluster or by a mutated or synthetic segment. Vectors and
CC host cells, and methods for producing novel polyketides by
CC culturing host cells are claimed. The polyketides obtained are
CC useful as antibiotics and insecticides. Fermentation products
CC containing C22-C23 dihydroavermectin, ivermectin and B1
CC avermectins are claimed.
XX
SQ Sequence 12381 BP; 1884 A; 4561 C; 4005 G; 1931 T; 0 other;

Query Match 100.0%; Score 534; DB 21; Length 12381;
Best Local Similarity 100.0%; Pred. No. 4.3e-90;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11426 GGGTCGGTGTTGGTGACGGGTGGGACGGGTGTGCTGGGTGCGGCGGTGGCGCGGCATCTG 11485

Qy         61 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11486 GCTGGTGTGTGTGGGGTGCGGGATCTGCTGTTGGTGAGCCGGCGTGGTCCGGATGCTCCG 11545

Qy        121 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11546 GGTGCGGAGGGTCTGCGGGCGGAGCTGGCCGCGTTGGGGGCGGAGGTGCGGATTGTTGCG 11605

Qy        181 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11606 TGTGATGTGGGGGAGCGGCGGGAGGTGGTCCGGCTGCTGGAGGGTGTTCCTGCCGGGTGT 11665

Qy        241 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11666 CCGCTGACGGGTGTCGTGCATGCGGCTGGTGTGCTGGACGATGCGACGATCGCCTCTCTC 11725

Qy        301 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      11726 ACGCCCGAGCGGCTGGGCACGGTGTTCGCGGCCAAGGTGGATGCCGCTCTTTTGCTGGAT 11785

Qy        361 GAGCTGACGCGGGGTATGGAGCTGTCGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATC 420
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2003, 11:26:45 ; Search time 106.227 Seconds
(without alignments)
11320.736 Million cell updates/sec

Title: US-09-914-286-1_COPY_10609_11142
Perfect score: 534
Sequence: 1 gggtcggtgttggtgacggg.....cgctggcgtgggggctgtgg 534

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	534	100.0	11916	22	AAH79279	Streptomyces averm
2	534	100.0	12381	21	AAZ58381	Streptomyces averm
3	534	100.0	30690	21	AAA92301	S. avermitilis ave
4	534	100.0	30690	22	AAH79277	Streptomyces averm
5	532.4	99.7	31422	21	AAA92302	S. avermitilis ave
6	532.4	99.7	31422	22	AAH79278	Streptomyces averm
7	257.2	48.2	2700	22	AAD17193	ERD48 insert DNA i
8	257.2	48.2	65140	22	AAD17184	Streptomyces nours
9	257.2	48.2	125401	22	AAD17186	Streptomyces nours
10	250.8	47.0	43280	18	AAT80413	Tylactone synthase
c 11	250.8	47.0	4411529	22	AAI99682	Mycobacterium tube
c 12	249.2	46.7	4403765	22	AAI99683	Mycobacterium tube
13	244	45.7	27541	22	AAD17185	Streptomyces nours
14	239.6	44.9	16767	22	AAF88339	S. spinosa DNA fra
15	239.6	44.9	29736	22	AAF88317	S. spinosa DNA fra
16	239.6	44.9	50000	22	AAF88313	S. spinosa DNA fra
17	239.6	44.9	80161	20	AAZ21501	DNA fragment of Sa
18	230	43.1	14775	22	AAF88338	S. spinosa DNA fra
19	230	43.1	44377	18	AAT78508	Platenolide syntha
20	230	43.1	44377	18	AAT80414	Platenolide syntha
21	230	43.1	50000	22	AAF88316	S. spinosa DNA fra
22	223.2	41.8	20394	22	AAF24892	Pimaricin biosynth
23	220.2	41.2	33529	23	AAS17367	DNA sequence of S.
24	218.8	41.0	4403765	22	AAI99683	Mycobacterium tube
25	218.8	41.0	4411529	22	AAI99682	Mycobacterium tube
26	212.4	39.8	15872	21	AAZ87283	S. venezuelae vep
27	209.2	39.2	15872	18	AAT68715	Streptomyces venez
28	207.4	38.8	4851	22	AAH52062	Mycobacterium tube
29	200.6	37.6	53789	19	AAV21187	Amycolatopsis medi
30	199.6	37.4	13842	21	AAZ87297	S. venezuelae macr
31	199.6	37.4	36778	21	AAZ87318	S. venezuelae pik
32	199.6	37.4	37948	21	AAZ87285	S. venezuelae pik
33	199.6	37.4	38506	21	AAA75633	Nucleotide sequenc
34	199.6	37.4	38506	21	AAZ56001	Recombinant cosmid
35	199.6	37.4	47981	22	AAF30757	Micromonospora meg
36	198	37.1	13987	18	AAT80415	Hybrid srmG/tylG O
37	197.6	37.0	5676	19	AAV21186	Amycolatopsis medi
38	196.4	36.8	11220	21	AAZ87298	S. venezuelae macr
39	194.8	36.5	9513	22	AAF88337	S. spinosa DNA fra
40	192	36.0	29879	14	AAQ46806	eryA region of S.
41	183.6	34.4	68750	21	AAZ55887	Sorangium cellulos
42	182.4	34.2	6459	22	AAF88336	S. spinosa DNA fra
43	182	34.1	71989	21	AAA29349	Sorangium cellulos
44	178.8	33.5	28598	17	AAT06769	Sorangium cellulos
45	178.8	33.5	28958	18	AAT89956	Sorangium cellulos

ALIGNMENTS

RESULT 1
AAH79279
ID AAH79279 standard; DNA; 11916 BP.
XX
AC AAH79279;
XX
DT 04-DEC-2001 (first entry)
XX
DE Streptomyces avermitilis coding sequence derivative SEQ ID NO: 3.
XX
KW Avermectin aglycone synthase; AAS; avermectin derivative;
KW drug production; vetinary drug; pesticide; ds.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..11916
FT /*tag= a
FT /product= "AAG65268"
FT /partial
XX
PN WO200162939-A1.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-JP01381.
XX
PR 24-FEB-2000; 2000JP-0047405.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
PA (KITA) KITASATO INST.
XX

THIS PAGE BLANK (USPTO)

the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."

BASE COUNT 24 a 13 c 838 g 130 t
ORIGIN

Query Match 1.0%; Score 295.2; DB 17; Length 1005;
Best Local Similarity 59.0%; Pred. No. 1e-31;
Matches 525; Conservative 0; Mismatches 363; Indels 2; Gaps 1;

```
Qy      18249 CACCCTCCCCACCACCCACCAACCCCACCAAACCTGGCTCATCGCCATCCCCGAAACCCA 18308
              | ||| | |||| |||  |  |||| ||   ||   | |  | ||  ||||    |||
Db       1003 CCCCCCACACACCCCCCCACCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCC 944

Qy      18309 GACCCACCACCCCCACATCACCAACATCCTCACCAACCTCCACCACCACGGCATCACCCC 18368
               ||| |||||||| |  | ||  |   | | ||  || || || || |  |  | ||||
Db        943 CCCCCCCCACCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCACCACCCCCC 884

Qy      18369 CATCCCCCTCACCCTCAACCACACCCACACCAACCCCCAACACCTCCACCACACCCGACA 18428
              || ||||| ||||| |  |||| ||| | ||  |||||  | || || || | |||  |
Db        883 CACCCCCCCCACCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 824

Qy      18429 ACAAGCCCAAAACCACACCACCGGACCCATCACCGGCCTGCTCTCCCTCCTCGCCCTCGA 18488
              ||   |||    || | || ||   |||  | ||  ||  | | ||| || | ||| |
Db        823 ACCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 764

Qy      18489 CGAAACACCCCACCCCCACCACCCCCACACACCCACCGGCACCCTCCTCAACCTCACCCT 18548
              |    | | || ||||| || ||||| | | ||| ||  | ||| || |  || | |||
Db        763 CCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 704

Qy      18549 CACCCAAACCCACACCCAAACCCACCCACCAACCCCCCTCTGGTACGCCACCACCAACGC 18608
              | |||   ||| | |||   ||| ||| ||  |||||| |     | || || ||  | |
Db        703 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 644

Qy      18609 CACCACCACCCACCCCAACGACCCCCTCACACACCCCACCCAAGCCCAAACCTGGGGACT 18668
              | || || |||||||  |  ||||| | | | |||| |||   | |   ||       |
Db        643 CCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCCCC 584

Qy      18669 CGCCCGCACCACCCTCCTCGAACACCCCACCCACACCGCCGGAATCATCGACCTCCCCAC 18728
              | ||| | || ||| || |    | |||| ||| |||| ||     |  |  || |||| |
Db        583 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCC 524

Qy      18729 CACCCCCACCCCCCACACCCTCCACCACCTCACCCAAACCCTCACCCAACCCCA--CCAC 18786
              | ||||| |||||| | ||| || ||| | | ||| |||||  | |||  |||||  || |
Db        523 CCCCCCCCCCCCCCCCCCCCCCCCCCCCAACACCCCCCAACCCCCCCCCCCCCCCACCCCCC 464

Qy      18787 CAAACCCAACTCGCCATCCGCACCACCGGCACCCACACCCGCCGCCTCACCCCCACCACC 18846
              |   |||  | | ||  |  | |||||  | ||| | |||  | || | |||||  ||  |
Db        463 CCCCCCCCCCACCCCCACACCCCCACCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCAC 404

Qy      18847 CTCACCCCCACACACCAACCACCCACCCCCACCCCCCACGGAACCACCCTCATCACCGGC 18906
              |    |||| |    ||  || ||| || || ||||||  |     || | | |  | ||  |
Db        403 CAACACCCCCCCTCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCACCCCCCCCCCC 344

Qy      18907 GGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCACCCACCAACCCACCCAA 18966
                 |||  | ||| | || ||| || | ||| ||||| | || ||| || |||| ||||
Db        343 CCCACCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCACCCCCCAC 284

Qy      18967 CACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACGCACAACACCTCACCACC 19026
              | || || |   ||   || ||    |||||||| |||||| | | |  | || | ||| |
Db        283 CCCCCCCCCAAAACACCCCAAAAACCCCCCCCACCCCCCCCCCCCCCCCCCCCCACCCCAAC 224

Qy      19027 CAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCGACACCAGCAACCCAGAC 19086
              || |   ||      |  || || || ||||  | ||  |  | | || | || |||   |
Db        223 CACCCAAAAACCCCAACCACCCCCCCCACCCCCCCCCGCGCCAAAACAACCACCCCCACC 164

Qy      19087 CAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCCTCACCAC 19136
               ||   |  |  | ||    ||||  ||||    |  || ||| | || |
Db        163 AAAAAACCCCCCCACCAACCCACCCCCCCCTGCGAGTACGCCCCCCCCC 114
```

Search completed: June 19, 2003, 02:46:12
Job time : 24152 secs

Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates , Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

BASE COUNT 85 a 19 c 930 g 96 t 36 others
ORIGIN

Query Match 1.0%; Score 295.6; DB 12; Length 1166;
Best Local Similarity 53.1%; Pred. No. 8.9e-32;
Matches 607; Conservative 0; Mismatches 534; Indels 2; Gaps 1;

```
Qy       4496 CCCTCCCCACCACCCACCAACCCCACCAAACCTGGCTCATCGCCATCCCCGAAACCCAGA 4555
             ||||| || ||     ||  |  || | |  ||  |||  | |   | |     | |
Db       1147 CCCTCTCCGCCTATTCCCTCCANCAACTACTCTCCCTCCCCCCTCCCGCACTCCCTCCAT 1088

Qy       4556 CCCACCACCCCCACATCACCAACATCCTCACCAACCTCCACCACCACGGCATCACCCCCA 4615
             |||  | ||||     | |||    || | ||  || ||  | || |  |  | | |||
Db       1087 CCCCTCCCCCCTCACACCCCACTCCCCCCCCCCCCCCCCCACCCCCCCCCTCCCCTCCCC 1028

Qy       4616 TCCCCCTCACCCTCAACCACACCCACACCAACCCCCAACACCTCCACCACACCCTCCACC 4675
             ||  |||||||| || |  | ||||| | ||||||    | | ||  | | |||||| |
Db       1027 TCTGCCTCACCCCCATCATCTCCCACCCAAACCCCACCCCTCCCCTACCCCCCCTCCCCA 968

Qy       4676 ACACCCGACAACAAGCCCAAAACCACACCACCGGAGCCATCACCGGCCTGCTCTCCCTCC 4735
               | |||    |  | |||    ||   || ||    ||  | |||   |  | | ||| ||
Db        967 CCTCCCCCACAGCACCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCC 908

Qy       4736 TCGCCCTCGACGAAACACCCCACCCCCACCACCCCCACACACCCACCGGCACCCTCCTCA 4795
              | ||| |  |    | |||| | ||| || ||||| | |  | ||| | | | || |
Db        907 CCCCCCCCTCCCCCCCCCCCCCCGCCCCCCCCCCCCCCCCCCTACAACCCCCCCTCCCCCCC 848

Qy       4796 ACCTCACCCTCACCCAAACCCACACCCAAACCCACCCACCAACCCCCCTCTGGTACGCCA 4855
              ||| | |||||  ||  |    | | ||  ||| | | || | ||||  |     | |||
Db        847 ACCCCCCCCTCCACCCTATNTCCCCACACCCCCCCTCCCCCATCCCCACCCCCCCCCCCA 788

Qy       4856 CCACCAACGCCACCACCACCCACCCCAACGACCCCCTCACACACCCCACCCAAGCCCAAA 4915
              ||  ||    | |   ||||| |||| |  |||||  | | | | ||||||   |||
Db        787 CCCACATATTCCCACACACCCCCCCC--CCCCCCCCGCCCCCCCTCCACCCCCCCCCCCC 730

Qy       4916 CCTGGGGACTCGCCCGCACCACCCTCCTCGAACACCCCACCCACACCGCCGGAATCATCG 4975
              ||      | | ||| | ||||||| || | ||| |||| ||| | || ||     |  |
Db        729 CCCCCCCCCCCCCCCCCCCCCACCCGCCCCCAACCCCCCCCCCCCCCCCCCCCCCCCCCC 670

Qy       4976 ACCTCCCCACCACCCCCACCCCCCACACCCTCCAGCACCTCACCCAAACCCTCACCCAAC 5035
              || |||| || |||||  ||||| | ||| ||  | || | |||   ||| | |||  |
Db        669 CCCCCCCCCCCCCCCCCCCACCCCCTCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 610

Qy       5036 CCCACCACCAAACCCAACTCGCCATCCGCACCACCGGCACCCACACCCGCCGCCTCACCC 5095
              ||| || ||  |||| | | ||  || | || ||  | |||   ||| |  || | |||
Db        609 CCCCCCCCCCCACCCCCCCCCCCCACCCCCCCCCCCCCCCCCCACCCCCCACCCCCCCCC 550

Qy       5096 CCACCACCCTCACCCCCACACACCAACCACCCACCCCCACCCCCCACGGAACCACCCTCA 5155
              || || ||| | ||||| |   ||  || ||| ||||| ||||||  |    || ||| |
Db        549 CCCCCCCCCCCCCCCCCCCCCATCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 490

Qy       5156 TCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCACCCACCAAC 5215
              | ||  |    ||  | |||   || ||| || | ||  || || |||| ||||||
Db        489 CCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCACCCCCCCCACCCCCCACCCCA 430

Qy       5216 CCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACGCACAACACC 5275
              || |||  | || || || | |||    ||    ||| | |  ||||| | | |  | |
Db        429 CCCCCCCCCCCCCCCCCCCCCCCCCACAACAAAACNCCCNCCCCCNCCCCCNCCCCCCCCCN 370

Qy       5276 TCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCGACACCAGCA 5335
              | || |||  |  ||| ||      | ||| || | ||  | || ||  |  |||   |
Db        369 CCCCCCCCCCTCCACAAAAACTCCCCTTCCTCCCCCCCCCCCCCCCCCCCCCCCACACACC 310

Qy       5336 ACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCCTCACCACCG 5395
              | |   ||  | ||  |  |       | ||  |||||| |  | |||||  | || ||
Db        309 CCTCCCCCCCCCCCCCCCCCCCCCNNNNNCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCC 250

Qy       5396 TCATCCACACCGCAGGCATCCTCGACGACGCCCACCCTCACCAACCTCACCCCCACCCAAC 5455
              |  ||   || |       ||    |  |      || |  |  || | ||||| |||  |
Db        249 CCCCCCCNCCCCCCCCCNCCCCCNCCCCCCCCNNCTCCNCNNCCNCCNCCCCCCCCCCCCC 190

Qy       5456 TCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAACTCACCCAAC 5515
              |      || || | |    ||| |  |||  | | |||||  |||  | |||  |
Db        189 CNNCCCNNCCCCCCCCCCCNNCCCCCCCCCCNCGCAAAACACCTCCCNCAAAACCCCCCCC 130

Qy       5516 ACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCGGCGCACCCG 5575
              | |||||  ||  |    || || || | |  |  |        |  | | |||
Db        129 NCCCCCCCAAAACTCCTCCTCCCCCCTCCTTCCCTCTCANGGTTTTTAAACCCCCCCCCT 70

Qy       5576 GCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACCACCGCCACA 5635
              |   |  |||     || || | | |  |       |  |  |  || || || |
Db         69 CACCCCCGCCCTAGTTTNCACCCCAGGGCNCCGCGCCGCCAACCCTCTCCCCCCCCCGCC 10

Qy       5636 CCC 5638
              |||
Db          9 CCC 7
```

RESULT 15
BH148610/c
LOCUS BH148610 1005 bp DNA linear GSS 27-AUG-2001
DEFINITION ENTPZ67TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION BH148610
VERSION BH148610.1 GI:15308098
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 1005)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 501
High quality sequence stop: 604.
FEATURES Location/Qualifiers
source 1. .1005
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for

(http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT 121 a 16 c 991 g 96 t 27 others
ORIGIN

Query Match 1.0%; Score 297.6; DB 10; Length 1251;
Best Local Similarity 52.6%; Pred. No. 4.7e-32;
Matches 641; Conservative 0; Mismatches 574; Indels 4; Gaps 1;

```
Qy       4446 CCAAGCCCGCATCAACACCTGGACCTACCAGGAAACCTGGAAACCCCTCACCCTCCCCAC 4505
Db       1248 CCTCCCCCACCTCTCCCCCTCCCTCAACCTGTCTTACCCCCCCCCCCCACGCCACCCTTA 1189

Qy       4506 CACCCACCAACCCCACCAAACCTGGCTCATCGCCATCCCCGAAACCCAGACCCACCACCC 4565
Db       1188 TGCATCCCTCCCCCCCCCCTCCTCCATGCCCCCTCACACAACAACCCACTCCCTTTTCCC 1129

Qy       4566 CCACATCACCAACATCCTCACCAACCTCCACCACCACGGCATCACCCCCATCCCCCTCAC 4625
Db       1128 CCCTCCTCATCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCTCCTCCCCCCCACTCCCCCT 1069

Qy       4626 CCTCAACCACACCCACACCAACCCCCAACACCTCCACCACACCCTCCACCACACCCGACA 4685
Db       1068 CCACTCNCCCACCCCCCCCCCCACCCCCCCCCTCTCCCCCCCCCCCTCCCTTCCTAATCC 1009

Qy       4686 ACAAGCCCAAAACCACACCACCGGAGCCATCACCGGCCTGCTCTCCCTCCTCGCCCTCGA 4745
Db       1008 CTCTCCCCCCCACCTCCCCCCCCTACCCACTTTCCCCCACTCCTCCAGCCCCCCCCTCCA 949

Qy       4746 CGAAACACCCCACCCCCACCACCCCCACACACCCACCGGCACCCTCCTCAACCTCACCCT 4805
Db        948 CCATCCCCCCCCCCCCCCCCCCCCCCCTCCCCTCCCCCCCCCACACCCCCCCCCTCCCCCC 889

Qy       4806 CACCCAAACCCACACCCAAACCCACCCACCAACCCCCCTCTGGTACGCCACCACCAACGC 4865
Db        888 CCCCCAACCCCCCCACTCCCCCCCCCCCCCCCCCCCCCCCC----CTCCCCTCCCTCCCT 833

Qy       4866 CACCACCACCCACCCCAACGACCCCCTCACACACCCCACCCAAGCCCAAACCTGGGGACT 4925
Db        832 CCCCCCCCCCCACCCCCACCCCCCCCCCCTCCCCCCCCCCCCCCTCCCCGCCCCCCCCCCG 773

Qy       4926 CGCCCGCACCACCCTCCTCGAACACCCCACCCACACCGCCGGAATCATCGACCTCCCCAC 4985
Db        772 CCCCCCCCCCACCCACTCCCCCCCCTCCCCCCACCCCCCCCCCCTCCCCGCACCCCCCCCC 713

Qy       4986 CACCCCCACCCCCCACACCCTCCAGCACCTCACCCAAACCCTCACCCAACCCCACCACCA 5045
Db        712 CCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCTCCNCCCCCCCCCCCCCCCCC 653

Qy       5046 AACCCAACTCGCCATCCGCACCACCGGCACCCACACCCGCCGCCTCACCCCCACCACCCT 5105
Db        652 CCTCCCACCGCCTCTCACCCCCCCCCCCCCCCCACCCCCCCCACCCCCCCCCCCCCCCCC 593

Qy       5106 CACCCCCACACACCAACCACCCACCCCCACCCCCCACGGAACCACCCTCATCACCGGCGG 5165
Db        592 CCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCACCCCCCCCCCCCCCCCCCCCCCTC 533

Qy       5166 AACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCACCCACCAACCCACCCAACA 5225
Db        532 CCCCCCCCCGCCCCCCTCCCCCCCCCCCCCCCCCACACCCCCCCCCCCCCCCCCCCCCCC 473

Qy       5226 CCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACGCACAACACCTCACCACCCA 5285
Db        472 CCCCCCCCCNTCCACACCCCCCCCCCCCNCCCCCCCCCCCCCCCTCNTNCNCNNCCCCCCC 413

Qy       5286 ACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCGACACCAGCAACCCAGACCA 5345
Db        412 CCCACAAAAAAAACTCCCCCCTCCCCCCTCCCCCCCCCCCTACCTCCCTCCCCCCCCACCC 353

Qy       5346 ACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCCTCACCACCGTCATCCACAC 5405
Db        352 TCCCCCCCCCCCCCCCCCCCCCCCCCCCCNNCCCCCCCCCCCCNNNCCCNTCNCCCCCCCCC 293

Qy       5406 CGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCCCCACCCAACTCAACAACGT 5465
```

```
Db        292 CCCCCCCCCCCCCCACCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 233

Qy       5466 CCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAACTCACCCAACACACCCCCCT 5525
Db        232 CCCCCCCCCCCCCCCNNCCCCCCCCCCCNACCGACCCTCTCTCCTCTAAAAAAACCCCCCCC 173

Qy       5526 CACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCGGCGCACCCGGCCAAGCCAA 5585
Db        172 CCCTAAAAAAACCCCCCCTCCCCCTCCCCCTTCGCCCCGCNNGTTNCTCCCCACACCCCT 113

Qy       5586 CTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACCACCGCCACACCCACCACCT 5645
Db        112 NCCCCCCCAAAACCACNTGTTTTAAATATCTTCCCCCCCCCCACCGNCCCCCCCCCCTCTC 53

Qy       5646 CCCCGCCACCAGCATCGCC 5664
Db          52 CCTCCCCCTCCCTTTCTCC 34
```

RESULT 14
BF256751/c
LOCUS BF256751 1166 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMEf0010O11f Hordeum vulgare seedling root EST library HVcDNA0007 (Etiolated and unstressed) Hordeum vulgare cDNA clone HVSMEf0010O11f, mRNA sequence.
ACCESSION BF256751
VERSION BF256751.2 GI:16315527
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 1166)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton,R.D., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex unstressed seedling root cDNA library
JOURNAL Unpublished (2001)
COMMENT On Nov 16, 2000 this sequence version replaced gi:11185864.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 172
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence start: 16
High quality sequence stop: 1043.
FEATURES Location/Qualifiers
source 1. .1166
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEf0010O11f"
/clone_lib="Hordeum vulgare seedling root EST library HVcDNA0007 (Etiolated and unstressed)"
/tissue_type="Seedling root"
/lab_host="TJC121"
/note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1; Seeds were surface sterilized then germinated under axenic conditions in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered crystallization dishes. Five-day old seedling roots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids. These steps were performed in the TJ Close laboratory at the University of California,

```
BASE COUNT       89 a     15 c    937 g     33 t      8 others
ORIGIN

  Query Match             1.0%;  Score 297.6;  DB 10;  Length 1082;
  Best Local Similarity   54.8%;  Pred. No. 4.7e-32;
  Matches  579;  Conservative    0;  Mismatches  477;  Indels    0;  Gaps    0;

Qy      18244 CCCCTCACCCTCCCCACCACCCACCAACCCCACCAAACCTGGCTCATCGCCATCCCCGAA 18303
              ||||||  | |||||  ||  ||   ||   ||   |    | |  | |  ||||
Db       1056 CCCCTCCTCTTCCCCCCCCCCTTCCGCTCCATCCCCTCTCCTCCCCCCTCTCCCCCCACC 997

Qy      18304 ACCCAGACCCACCACCCCCACATCACCAACATCCTCACCAACCTCCACCACCACGGCATC 18363
                ||   |||    ||||| |  | ||  |  |||| ||  ||  | || || |  |  |
Db        996 CTCCCACCCCCNACCCCCCCCCCCCCCCCCCTCCCTCGCCTCCCCGCCCCCCCTCCCCCCC 937

Qy      18364 ACCCCCATCCCCCTCACCCTCAACCACACCCACACCAACCCCCAACACCTCCACCACACC 18423
              |||||  |||||    || ||  || | ||| | ||  |||||  | ||||| ||    ||
Db        936 CCCCCCCCCCCCCCTCCCGTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCTCCC 877

Qy      18424 CGACAACAAGCCCAAAACCACACCACCGGACCCATCACCGGCCTGCTCTCCCTCCTCGCC 18483
              |     |   |||     || |  | ||   |||     ||  ||| | | |||||| | ||
Db        876 CCCTCCCCCCCCCTCCTCCCCCTCTCCCCCCCCCGTCCCCCCCTCCCCCCCCTCCCCCCC 817

Qy      18484 CTCGACGAAACACCCCACCCCCACCACCCCCACACACCCACCGGCACCCTCCTCAACCTC 18543
              | |  ||   | |||| ||||| || || |  | | | |  |   |  | || |  || |
Db        816 CCCCTCGCCCCCCCCCCCCCCCCCCCCCCTCTCCCCCCTCTTCCCCCTTCCCCCCCCCCC 757

Qy      18544 ACCCTCACCCAAACCCACACCCAAACCCACCCACCAACCCCCCTCTGGTACGCCACCACC 18603
              |||| | |||   ||| | |||   ||| ||| ||  ||||||||      | || || ||
Db        756 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCC 697

Qy      18604 AACGCCACCACCACCCACCCCAACGACCCCCTCACACACCCCACCCAAGCCCAAACCTGG 18663
                 || |  || ||| |||| |  ||||| | | | |||| |||   |||   ||
Db        696 CCTCCCCCTCCCCCCCCCCCCTTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 637

Qy      18664 GGACTCGCCCGCACCACCCTCCTCGAACACCCCACCCACACCGCCGGAATCATCGACCTC 18723
                | ||| | || | |||| |   | |||| ||| | || ||     |  |  || |
Db        636 TCCTCCTCCCCCCCCCCTCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 577

Qy      18724 CCCACCACCCCCACCCCCCACACCCTCCACCACCTCACCCAAACCCTCACCCAACCCCAC 18783
              ||| || ||||| ||||||  | ||| || ||  | | |||  |||| | ||  ||||
Db        576 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCACCCCCCCCCCCCCCCC 517

Qy      18784 CACCAAACCCAACTCGCCATCCGCACCACCGGCACCCACACCCGCCGCCTCACCCCCACC 18843
              | ||   |||  | | ||  || | || ||  | ||| | ||| || || | |||||  |
Db        516 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTC 457

Qy      18844 ACCCTCACCCCCACACACCAACCACCCACCCCCACCCCCCACGGAACCACCCTCATCACC 18903
                || | ||||| | | ||  || ||| ||||  |||||  |    || ||| |  | ||
Db        456 TTCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCC 397

Qy      18904 GGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCACCCACCAACCCACC 18963
                |    ||  | ||| | || ||| || | ||| || || | || ||| ||  ||| ||
Db        396 TCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 337

Qy      18964 CAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACGCACAACACCTCACC 19023
              || |   || || | ||  ||   ||  ||||| | |||||| | | |  | || | ||
Db        336 AAAAAAACCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCC 277

Qy      19024 ACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCGACACCAGCAACCCA 19083
              |||  | ||  |       |  || || | ||  | || ||  |  |||   || ||
Db        276 CCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCACACCCACCCAC 217

Qy      19084 GACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCCTCACCACCGTCATC 19143
                || | || ||| | || |  | ||  |||||| |  | ||||| |  | |||  |  |
Db        216 TCCCCCCCCCNACATCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCNCC 157

Qy      19144 CACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCCCCACCCAACTCAAC 19203
              | | ||    |  ||      |  || ||| | ||        ||||| |||  | |  |
Db        156 CCCCCCCGCCCCTCCCCTCCTAAAACCCCCCCCCCCCCCAAAANCCCCCCCCCCCCCCCC 97

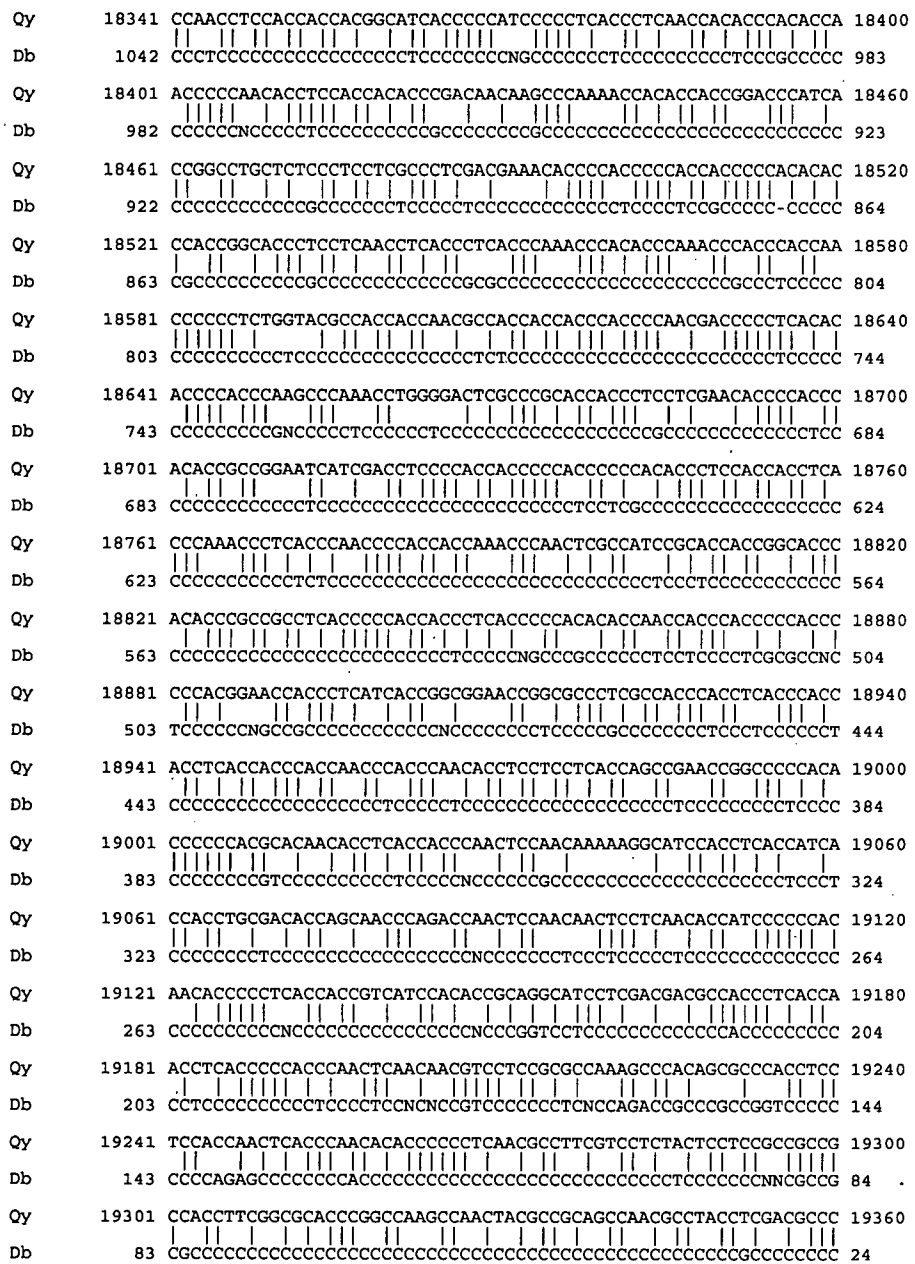
Qy      19204 AACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAACTCACCCAACACACC 19263
                |   |      ||  ||| |  |  | | |     | . | | ||   |  |
```

```
Db         96 TGGGCGNTGTCTTTCACCNCCCCCCCCCCAACCCNTGTGTAATTTCCCCCCCACGCCTTCT 37

Qy      19264 CCCCTCAACGCCTTCGTCCTCTACTCCTCCGCCGCC 19299
              |||| |  | |||||  ||  | |  | || || ||
Db         36 CCCCCCCCCCCCCTTCTCCCCTTTCCTCCCCCCCCCC 1


RESULT 13
BE455089/c
LOCUS       BE455089               1251 bp    mRNA    linear   EST 22-OCT-2001
DEFINITION  HVSMEh0096C10f Hordeum vulgare 5-45 DAP spike EST library
            HVcDNA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMEh0096C10f,
            mRNA sequence.
ACCESSION   BE455089
VERSION     BE455089.1  GI:9464414
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 1251)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton
            ,R.D., Close,S.J., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Morex 5-45 DAP spike cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 157
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 115
            High quality sequence stop: 1233.
FEATURES             Location/Qualifiers
     source          1. .1251
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEh0096C10f"
                     /clone_lib="Hordeum vulgare 5-45 DAP spike EST library
                     HVcDNA0009 (5 to 45 DAP)"
                     /tissue_type="5-45 DAP Spike"
                     /lab_host="SOLR"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1;
                     Plants were grown in the greenhouse at the University of
                     California, Riverside (Fenton, SJ Close, TJ Close). Whole
                     spikes with awns trimmed were collected at 5, 10, 15, 20,
                     30 and 45 DAP (Fenton). Total RNA was prepared from each
                     pool, equal quantities of all six RNA pools were combined,
                     poly(A) RNA was purified from the mixture, one primary
                     unamplified cDNA library was made, and 1 million pfu were
                     in vivo excised to give pBluescript SK(-) cDNA phagemids
                     (Choi)in the TJ Close lab at the University of California,
                     Riverside. Phagemids were plated and picked at the Clemson
                     University Genomics Institute (CUGI) (Begum, Palmer,
                     Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
                     sequencing and sequence analysis were performed at CUGI
                     (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
                     The sequence has been trimmed to remove vector sequence
                     and contains a minimum of 100 bases of phred value 20 or
                     above. For more details on library preparation and
                     sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
```

Query Match 1.0%; Score 298.2; DB 13; Length 1042;
Best Local Similarity 55.4%; Pred. No. 3.9e-32;
Matches 578; Conservative 0; Mismatches 464; Indels 1; Gaps 1;



Qy 19361 TCGCCCACCACCGCCACACCCAC 19383
Db 23 CCCCCCCCCCCCCCCCCCCCCCC 1

RESULT 12
BE455154/c
LOCUS BE455154 1082 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMEh0096G08f Hordeum vulgare 5-45 DAP spike EST library HVcDNA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMEh0096G08f, mRNA sequence.
ACCESSION BE455154
VERSION BE455154.2 GI:16322259
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 1082)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton ,R.D., Close,S.J., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex 5-45 DAP spike cDNA library
JOURNAL Unpublished (2001)
COMMENT On Jul 26, 2000 this sequence version replaced gi:9464545.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 280
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence start: 3
High quality sequence stop: 1073.
FEATURES Location/Qualifiers
source 1. .1082
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEh0096G08f"
/clone_lib="Hordeum vulgare 5-45 DAP spike EST library HVcDNA0009 (5 to 45 DAP)"
/tissue_type="5-45 DAP Spike"
/lab_host="SOLR"
/note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1; Plants were grown in the greenhouse at the University of California, Riverside (Fenton, SJ Close, TJ Close). Whole spikes with awns trimmed were collected at 5, 10, 15, 20, 30 and 45 DAP (Fenton). Total RNA was prepared from each pool, equal quantities of all six RNA pools were combined, poly(A) RNA was purified from the mixture, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Choi)in the TJ Close lab at the University of California, Riverside. Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

crystallization dishes. Five-day old seedling roots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids. These steps were performed in the TJ Close laboratory at the University of California, Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates , Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

BASE COUNT 75 a 35 c 948 g 105 t 33 others
ORIGIN

Query Match 1.0%; Score 299.2; DB 12; Length 1196;
Best Local Similarity 53.8%; Pred. No. 2.8e-32;
Matches 607; Conservative 0; Mismatches 517; Indels 5; Gaps 1;

```
Qy      18243 ACCCCTCACCCTCCCCACCACCCACCAACCCCACCAAACCTGGCTCATCGCCATCCCCGA 18302
              |||||   ||  | || || |||||    ||||||   ||   | |  | ||  |||||
Db       1164 ACCCCCACCCGCCNCCCCCGCCCACTCCTCCCACCCCCCCCCACCCCCCCCCGCCCCCGC 1105

Qy      18303 AACCCAGACCCACCACCCCCACATCACCAACATCCTCACCAACCTCCACCACCACGGCAT 18362
                ||    || |||| ||| |  | |   |  || | ||  || |  || || |
Db       1104 CCACCCCCTCCCCCACACCCCCTCCTCGCCCGCCCCCCCCCCCCCACTCCCCCCCCCCTCC 1045

Qy      18363 CACCCCCATCCCCCTCACCCTCAACCACACCCACACCAACCCCCAACACCTCCACCACAC 18422
              | ||| || ||||| ||||| |  || |  ||   |||||| ||  | || |   |  |
Db       1044 CCCCCTCACCCCCCCCACCCCCCCCCCCCTCCCTCCCAACCGCCCCCTCCCCTCTCCAAA 985

Qy      18423 CCGACAACAAGCCCAAAACCACACCACCGGACCCATCACCGGCCTGCTCTCCCTCCTCGC 18482
              ||  | |||  | |         | || ||     |   | ||  ||  | | ||||||
Db        984 CCCGCCACACCCGCCCCCGGCCTCCCCCCTCGCTCCCCCCCCCCCTCCCCCCCTCCACCT 925

Qy      18483 CCTCGACGAAACACCCCACCCCCACCACCCCCACACACCCACCGGCACCCTCCTCAACCT 18542
              || |  |    ||| || ||||  || ||     |   ||| ||    ||| || |  ||
Db        924 CCCCACCCCCCCACACCCCCCCGCCCCCCGTTCCCTCCCCCCCCCTCCCCCCCCCCCCCC 865

Qy      18543 CACCCTCACCCAAACCCACACCCAAACCCACCCACCAACCCCCCTCTGGTACGCCACCAC 18602
              | ||| | ||    ||| | |||  |||| | ||||||||||||||       | || || |
Db        864 CCCCCCCTCCAGCCCCCCCCCCCCCCACCCCCACACCAACCCCCCTACCCCCCTCCCCCCC 805

Qy      18603 CAACGCCACCACCACCCACCCCAACGACCCCCTCACACACCCCACCCAAGCCCAAACCTG 18662
              || | || ||  |  |||| ||  | |||| | ||||| |||| |||   |||   ||
Db        804 CACCCCCCCCCACCACCACACCCCCCACCCACCCACACCCCCCCCCCCCCCCCCCCCCCA 745

Qy      18663 GGGACTCGCCCGCACCACCCTCCTCGAACACCCCACCCACACCGCCGGAATCATCGACCT 18722
                | | ||| | || ||| ||     | |||| ||| | || ||      |  |  ||
Db        744 CCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 685

Qy      18723 CCCCACCACCCCCACCCCCCACACCCTCCACCACCTCACCCAAACCCTCACCCAACCCCA 18782
              |||| || ||||| ||||   ||||  || || |      ||  ||| | |||  ||||
Db        684 CCCCCCCCCCCCCCCCCCCAACCACCACCCCCCCCTCACACCCCCCCCCCCCCCCCCCCC 625

Qy      18783 CCACCAAACCCAACTCGCCATCCGCACCACCGGCACCCACACCCGCCGCCTCACCCCCAC 18842
              || ||   |||  | | ||  || | ||  ||| | |||||  || || || |||||||
Db        624 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCACCACCCCCCCCCCACCCCCC 565

Qy      18843 CACCCTCACCCCCACACACCAACCACCCACCCCCACCCCCCACGGAACCACC-----CTC 18897
              | | | |||||  | | |  | | ||| ||||| || ||| |    || ||      | |
Db        564 CCCACCCCCCCCCCCCCCCGCATCCCCCCCCCCCCCCCGCCCCCACCCCCCCCACCCACCC 505

Qy      18898 ATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCACCCACCAA 18957
```

```
              | |  |  |      || |  || | ||||||  | | ||| || || | || ||| || |
Db        504 ACCGGCCCCCCCCCACGCCCACCCCCCCCACCCCACCCCCCCCCCCCCCCCCCCCCTCCCA 445

Qy      18958 CCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACGCACAACAC 19017
              ||  ||| ||||| |  |  | ||| ||   || | |||| | |||||     | |  | |
Db        444 CCACCCCCACACCNCACCNNCCCCCACCCCGCCCGGCCCCCCCCCCCCCTCTCCCCCCCNC 385

Qy      19018 CTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCGACACCAGC 19077
              | | || ||   | ||           |  || || | ||  | || ||  |  | ||
Db        384 CCCCCCCCCANCCCCTCNNCCCCCCCCCCCCCCCCCCCCCCCCCCGCCGCCCCCCCCAA 325

Qy      19078 AACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCCACAACACCCCCTCACCACC 19137
              ||      |  ||||     | || |  | ||| |||||| || |||| || |    ||
Db        324 AAAAATTTCTTTCTCCTCTCCCCCCCCCCCCCCCACCCCCCCCCACCACCACCACCNNCCC 265

Qy      19138 GTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCCCCACCCAA 19197
                | ||| |  ||  |  || |  |  | ||  || | ||  ||   ||||| ||||
Db        264 CCAACACACCCACCACCCCCCCCCCCCCNCNCCCTCCCCCCCCCCCCCACCCCCCCCCAC 205

Qy      19198 CTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAACTCACCCAA 19257
              | |  |  |  || |  |  | |  | | |     | | || || || || |    ||
Db        204 CCCCCCCCCCCCCCCNNCNNCCACNCNCCCCCNNCNCCCCCCCCCCCCCCCCCNNNCCCC 145

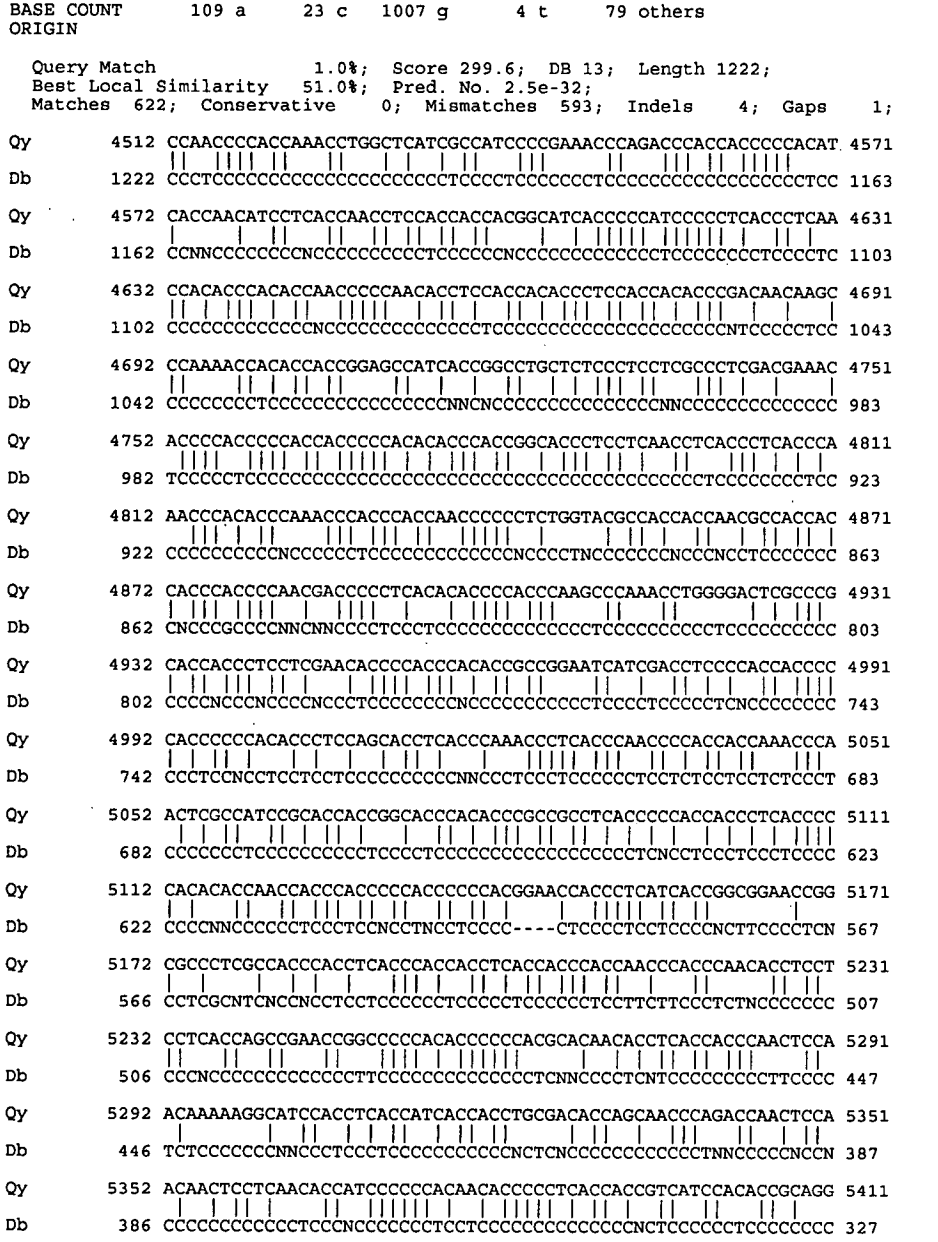
Qy      19258 CACACCCCCCTCAACGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCGGCGCACCC 19317
              | | |||||| | |  || ||     |   || ||     || || ||  |  | | ||
Db        144 CCCCCCCCCCCCGAACCCCTCCCAAACCNNCCCCCCCAAACCCCCCCCCCCCNCCCCCCT 85

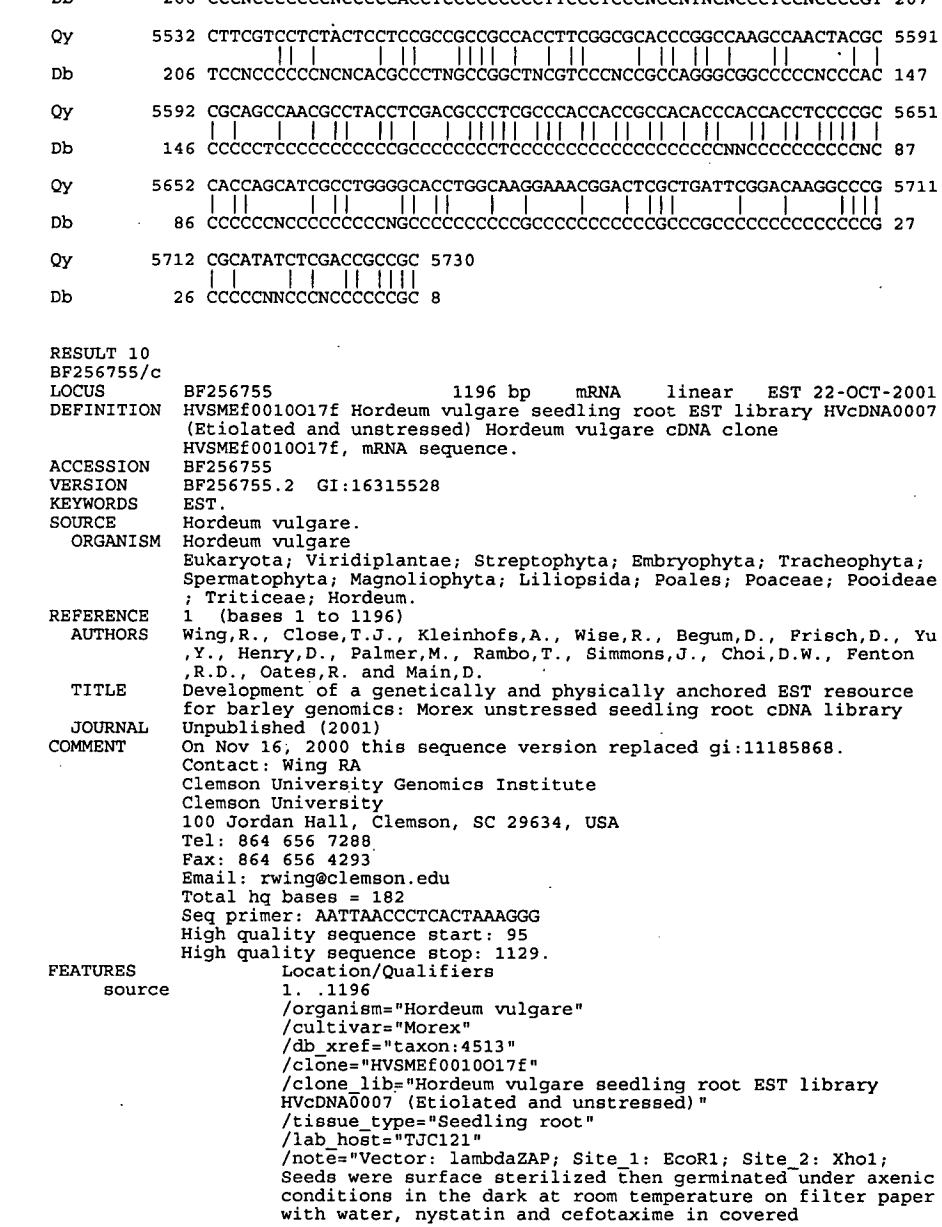
Qy      19318 GGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCC 19366
              |       ||||  | || |    |      ||| |     ||  |||||
Db         84 GAGGGCTTCAACCCCCCCCCCCAACGCNTGCTANATTCCACCCAGCGCCC 36
```

RESULT 11
BI416535/c
LOCUS BI416535 1042 bp mRNA linear EST 15-AUG-2001
DEFINITION hasp001xe11f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xe11f, mRNA sequence.
ACCESSION BI416535
VERSION BI416535.1 GI:15187558
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 1042)
AUTHORS Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026,S-750 07, Uppsala, Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES Location/Qualifiers
source 1. .1042
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="hasp001xe11f"
/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
/note="Vector: pT-Adv; Site_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidon annosum (FP5)."
BASE COUNT 54 a 37 c 929 g 6 t 16 others
ORIGIN

stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
/note="Vector: pT-Adv; Site_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidon annosum (FP5)."
BASE COUNT 109 a 23 c 1007 g 4 t 79 others
ORIGIN

Query Match 1.0%; Score 299.6; DB 13; Length 1222;
Best Local Similarity 51.0%; Pred. No. 2.5e-32;
Matches 622; Conservative 0; Mismatches 593; Indels 4; Gaps 1;





RESULT 10
BF256755/c
LOCUS BF256755 1196 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMEf0010O17f Hordeum vulgare seedling root EST library HVcDNA0007 (Etiolated and unstressed) Hordeum vulgare cDNA clone HVSMEf0010O17f, mRNA sequence.
ACCESSION BF256755
VERSION BF256755.2 GI:16315528
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 1196)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu ,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton ,R.D., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex unstressed seedling root cDNA library
JOURNAL Unpublished (2001)
COMMENT On Nov 16, 2000 this sequence version replaced gi:11185868.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 182
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence start: 95
High quality sequence stop: 1129.
FEATURES Location/Qualifiers
source 1. .1196
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEf0010O17f"
/clone_lib="Hordeum vulgare seedling root EST library HVcDNA0007 (Etiolated and unstressed)"
/tissue_type="Seedling root"
/lab_host="TJC121"
/note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1; Seeds were surface sterilized then germinated under axenic conditions in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1453)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 4
            Seq primer: Reverse.
FEATURES             Location/Qualifiers
     source          1. .1453
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGA"
                     /note="Vector: pBluescriptSKm"
BASE COUNT       23 a     31 c    901 g    382 t    116 others
ORIGIN

  Query Match             1.0%;  Score 301;  DB 10;  Length 1453;
  Best Local Similarity   53.6%;  Pred. No. 1.6e-32;
  Matches  661;  Conservative    0;  Mismatches  559;  Indels   13;  Gaps    5;

Qy       4429 CACCGCCACCAACACGACCAAGCCCGCATCAACACCTGGACCTACCAGGAAACCTGGAAA 4488
Db       1422 CACCACACANACCANCCACNACACCACCACAACANCACACCACACCACCACNACCCCACN 1363

Qy       4489 CCCCTCACCCTCCCCACCACCCACCAACCCCACCAAACCTGGCTCATCGCCATCCCCGAA 4548
Db       1362 CCCACCACACCACCCCACCCNACACNACNCACACACACAANACACACCAACNACACACAC 1303

Qy       4549 ACCCAGACCCACCACCCCCACATCACCAACATCCTCACCAACCTCCACCACCACGGCATC 4608
Db       1302 CCCCACACCCACACCCCACCACACACCCACCCACACCCACACCACCACCACC-CACCCAC 1244

Qy       4609 ACCCCCATCCCCCTCACCCTCAACCACACCCACACCAACCCCCAACACCTCCACCACACC 4668
Db       1243 ANCCCCACCNCCCNCNACCCACACCACACCCCCCCCACCCANCCACCACCACCCCCACCCC 1184

Qy       4669 CTCCACCACACCCGACAACAAGCCCAAAACCACACCACCGGAGCCATCACCGGCCTGCTC 4728
Db       1183 ACCACCACCCACNACAACCACAACCCCCACCACACACACNCAACNCACNACCACCACACC 1124

Qy       4729 TCCCTCCTCGCCCTCGACGAAACACCCCACCCCCACCACCCCCACACACCCACCGGCACC 4788
Db       1123 ACCCACACACCCCACNANCACCACCCACACACACCACCACCACACCCCAACACCCCCACC 1064

Qy       4789 CTCCTCAACCTCACCCTCACCCAAACCCACACCCAAACCCACCCACCAACCCCCCTCTGG 4848
Db       1063 ACCCCCACCACCNCCCACACCACCCCACCCCACCCCAACACACCACCCACCACNCCCCCA 1004

Qy       4849 TACGCCACCACCAACGCCACCACCACCCACCCCAACGACCCCCTCACACACCCCACCCAA 4908
Db       1003 CCANCCAACCCNCACCCACCCACACCCCACACCA---NCCCCCCCCCACNCCACNCCCAC 947

Qy       4909 GCCCAAACCTGGGGACTCGCCCGCACCACCCTCCTCGAACACCCCACCCACACCGCCGGA 4968
Db        946 CCCACACCCCACACCAACCCACCCACACCACCCCCCCACCAACCCACCCCANCACCC--A 889

Qy       4969 ATCATCGACCTCCCCACCACCCCCACCCCCCACACCCTCCAGCACCTCACCCAAACCCTC 5028
Db        888 ACCCACAACACCCACACCCCCCAACAACCCCCACCACCCCACCNCCACACCCANANCCCC 829

Qy       5029 ACCCAACCCCACCACCAAACCCAACTCGCCATCCGCACCACCGGCACCCACACCCGCCGC 5088
Db        828 ANCCCCCCCCCACCACCA----CACAACCACCACCCCANACCACACACCCCCNACACACCC 773

Qy       5089 CTCACCCCCACCACCCTCACCCCCACACACCAACCACCCACCCCCCACCCCCCACGGAACC 5148
Db        772 ACCACACCCACCCACACCCCCACCACACCCCACAACACACCCACACCACACCCCCCCACN 713

Qy       5149 ACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCACC 5208
Db        712 ACACNCACCACCACACACCCCAACCCACACACCCCNCCCAACCCCACACACACCACCACC 653

Qy       5209 CACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACGCA 5268
Db        652 CCACCACCCCCACCCCCCCCCCCCCCCACCACACCAACCCCCCCACCACCCACACCACCAC 593

Qy       5269 CAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCGAC 5328
Db        592 ACACCCCCACAACNCCACCACCCACCACACCCCACCAACCCACCCCAAACCACACACCCA 533

Qy       5329 ACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCCTC 5388
Db        532 CCCAACCCACCNCACCCAACCAACACACAACCACACACCCACCACCAACCCCACCCCCCC 473

Qy       5389 ACCACC---GTCATCCACACCGCAGGCATCCTCGACGACGCCCACCCTCACCAACCTCACC 5445
Db        472 ACCACCCAACCCCCCCACACCACCACCCCCCACCCANCCCCCACACCACACACCCACCCC 413

Qy       5446 CCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAA 5505
Db        412 CACACCCNACACACCACCCCCCCCACCCCCANACCACCCCACCCCCCCCCCCAACCCC 353

Qy       5506 CTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTC 5565
Db        352 CACCACNACCACCCCCCAACCACACCCCCANCCCCCCACCCCCCCCACCACCCCCANCCAC 293

Qy       5566 GGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCAC 5625
Db        292 ACCCAACCACNCCAACCGACCACCCCCCACACNCCNCCCCCCCNCACAACACCCCNCCCCC 233

Qy       5626 CACCGCCACACCCACCACCTCCCCGCCACCAGC 5658
Db        232 CCNNANCCNCCNCNCNCCNNCCCCCNCNCCCNC 200


RESULT 9
BI416660/c
LOCUS       BI416660               1222 bp    mRNA    linear   EST 15-AUG-2001
DEFINITION  hasp001xm19f Heterobasidion annosum - Scots pine infection stage
            subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
            annosum cDNA clone hasp001xm19f, mRNA sequence.
ACCESSION   BI416660
VERSION     BI416660.1  GI:15187683
KEYWORDS    EST.
SOURCE      Pinus sylvestris/Heterobasidion annosum.
  ORGANISM  Pinus sylvestris/Heterobasidion annosum
            Eukaryota; mixed EST libraries.
REFERENCE   1  (bases 1 to 1222)
  AUTHORS   Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
  TITLE     Expressed sequence tags of randomly selected cDNA clones from the
            interaction of the root rot fungus (Heterobasidion annosum) with
            seedling roots of Scots pine (Pinus sylvestris)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Fred O. Asiegbu
            Dept. of Forest Mycology & Pathology
            Swedish University of Agriculture, Box 7026,S-750 07, Uppsala,
            Sweden
            Tel: +46 18 67 15 98
            Fax: +46 18 30 92 45
            Email: Fred.Asiegbu@mykopat.slu.se
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .1222
                     /organism="Pinus sylvestris/Heterobasidion annosum"
                     /db_xref="taxon:169015"
                     /clone="hasp001xm19f"
                     /clone_lib="Heterobasidion annosum - Scots pine infection
```

```
RESULT 7
BH146899/c
LOCUS       BH146899               1222 bp    DNA     linear   GSS 27-AUG-2001
DEFINITION  ENTQM47TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   BH146899
VERSION     BH146899.1  GI:15302995
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 1222)
  AUTHORS   Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 452
            High quality sequence stop: 563.
FEATURES             Location/Qualifiers
     source          1. .1222
                     /organism="Entamoeba histolytica"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
BASE COUNT      125 a      0 c   1048 g     49 t
ORIGIN

  Query Match             1.0%;  Score 306.4;  DB 17;  Length 1222;
  Best Local Similarity   55.4%;  Pred. No. 2.8e-33;
  Matches  654;  Conservative    0;  Mismatches  521;  Indels    5;  Gaps    3;

Qy       4489 CCCCTCACCCTCCCCACCACCCACCAACCCCACCAAACCTGGCTCATCGCCATCCCCGAA 4548
Db       1179 CCCCCCCCCCCCCCCCCCCCCCTTCCCCCCACCCCCCCTTCCCCCCCCTCTCCCTTCTCC 1120

Qy       4549 ACCCAGACCCACCACCCCCACATCACCAACATCCTCACCAACCTCCACCACCACGGCATC 4608
Db       1119 CCCCCTCTCCCCCACCCCCCCCTCCCCTCCTTCCCCCCCCCCCTCCTCCCCCCCCCCCA 1060

Qy       4609 ACCCCCATCCCCCTCACCCTCAACCACACCCACACCAACCCCCAACACCTCCACCACACC 4668
Db       1059 CTCCCCTCCCCCTCTCCCCCTTCCCCCACCTCCCCCCCCCCCCACTCCTCCCCTCCCCCCC 1000

Qy       4669 CT--CCACCACACCCGACAACAAGCCCAAAACCACACCACCGGAGCCATCACCGGCCTGC 4726
Db        999 CTTCCCCCTTCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCACCCCCCCCCCCCCCCCCTCC 940

Qy       4727 TCTCCCTCCTCGCCCTCGACGAAACACCCCACCCCCACCACCCCCACACACCCACCGGCA 4786
Db        939 CCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 880

Qy       4787 CCCTCCTCAACCTCACCCTCACCCAAACCCACACCCAAACCCACCCACCAACCCCCCTCT 4846
Db        879 CCCCCCCCCCCCCCCCCTACCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCC 820

Qy       4847 GGTACGCCACCACCAACGCCACCACCACCCACCCCAACGACCCCCTCACACACCCCACCC 4906
Db        819 CCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCTCCCCCCCCCCCCCCC 760

Qy       4907 AAGCCCAAACCTGGGGACTCGCCCGCACCACCCTCCTCGAACACCCCACCCACACCGCCG 4966
Db        759 CTCCCCTTCCCTCCCCTCCTCCCCTCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCC 700

Qy       4967 GAATCATCGACCTCCCCACCACCCCCACCCCCCACACCCTCCAGCACCTCACCCAAACCC 5026
Db        699 CCCCCCTCCCCCTCCTCCCCTCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCACCCCCCCC 640

Qy       5027 TCACCCAACCCCACCACCAAACCCAACTCGCCATCCGCACCACCGGCACCCACACCCGCC 5086
Db        639 CCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCTC-CCCCCCCCCCCCCCTCCCCCC 581

Qy       5087 GCCTCACCCCCACCACCCTCACCCCCACACACCAACCACCCACCCCCACCCCCCACGGAA 5146
Db        580 CCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCC 521

Qy       5147 CCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCA 5206
Db        520 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 461

Qy       5207 CCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACG 5266
Db        460 CCCCCCCCCCCTCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCACCCCCCTCCCCCCCCCC 401

Qy       5267 CACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCG 5326
Db        400 TCTCCCTCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCTCCCCCCCCCCCCCCCCCCCC 341

Qy       5327 ACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCC 5386
Db        340 TTTTTCTCTTCCCCTCTCCCTCCCTTTTCCCCCCACAAACAACCCCCCCCCCCCCCCCC 281

Qy       5387 TCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCC 5446
Db        280 CCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 221

Qy       5447 CCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAAC 5506
Db        220 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCC 161

Qy       5507 TCACCCAACACACCCCCCTCACCGCCTTCGTCCTCT--ACTCCTCCGCCGCCGCCACCTT 5564
Db        160 CCCCCCCCCCCCCCCCTTATTCCCACCTTTTTTTTTTTCCCTCCCCCTCACCCACCACAAT 101

Qy       5565 CGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCA 5624
Db        100 TATTTTTTCCCCCCCTCCCCCTAAATCCCTAAAAACCCCCCCCAAAAACCCCCACCCCCA 41

Qy       5625 CCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCC 5664
Db         40 ACCACACCCCCCACCCCCCACACCCCACCCCCTTTTCACC 1


RESULT 8
AW949571/c
LOCUS       AW949571               1453 bp    mRNA    linear   EST 01-JUN-2000
DEFINITION  EST361641 MAGE resequences, MAGA Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW949571
VERSION     AW949571.1  GI:8139202
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Db        581 CCCCCCCCCCCACCCACCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 522

Qy      18844 ACCCTCACCCCCACACACCAACCACCCACCCCCACCCCCCACGGAACCACCCTCATCACC 18903
              ||| | ||| | |   ||  |||||  |  || || | |  |   || ||| |  |  |
Db        521 CCCCACCCCCACCCCACCCCCCCCACCACCGGCCCCCGCGCCGGCCCCCCCCCCCCCCCGC 462

Qy      18904 GGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCACCCACCAACCCACC 18963
              | |     || || ||| | || ||| || | ||| || || |     ||| |  ||| ||
Db        461 GCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGNNCCCCCGGCCCCCCC 402

Qy      18964 CAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACGCACAACACCTCACC 19023
              |  | || || || | ||  ||    ||  ||||| | |||||| | | |  | || | |
Db        401 CCCCCCCCCCCCCCCNCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCT 342

Qy      19024 ACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCGACACCAGCAACCCA 19083
              |||  | || ||||||     | | ||   || ||  |  |  |  | ||  || |||
Db        341 GCCCCCCCCCCACAAAATCTTTTTCCCCCTCCCCTCCTCCNCNCCCCCCCCCACACCCCC 282

Qy      19084 GACCAACTCCAACAACTCCTCAACACCATCCCCCCCACAACACCCCCTCACCACCGTCATC 19143
                ||  | ||  |  |     |  |      |||  | |  || |  | | || ||  |  |
Db        281 CCCCCCCCCCCCCCCCCNNNNCCCCNNNNNCCCNNC-CNNCANCNNCCCCCCCCCCCCCCC 223

Qy      19144 CACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCCCCACCCAACTCAAC 19203
              | | ||  |   | ||  |  | | || ||| | ||  || | ||||  |||  | |
Db        222 CCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNNCCCNCCCCCCN 163

Qy      19204 AACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAACTCACCCAACACACC 19263
              |  || ||||       |||| | | ||| || ||   | |  | | |||  | | ||
Db        162 CCCCCCCCCCGCAAACCCCCCCAAAACCCCCCCCCCCCTAAACCCCCCCCCCCCCCCCCC 103

Qy      19264 CCCCTCAACGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCGGCGCACCC 19317
              |||| |    | ||   |  |  | || |   || |     |       |||||
Db        102 CCCCCCNGGTCNTTNCACAACCNCNCCCCNAACGGCNNGCCTAANTNNCCACCC 49
```

RESULT 6
BI416575/c
LOCUS BI416575 931 bp mRNA linear EST 15-AUG-2001
DEFINITION hasp001xg15f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xg15f, mRNA sequence.
ACCESSION BI416575
VERSION BI416575.1 GI:15187598
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 931)
AUTHORS Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026,S-750 07, Uppsala, Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES Location/Qualifiers
source 1. .931
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="hasp001xg15f"
/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
/note="Vector: pT-Adv; Site_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidon annosum (FP5)."
BASE COUNT 9 a 7 c 884 g 15 t 16 others
ORIGIN

Query Match 1.0%; Score 310.4; DB 13; Length 931;
Best Local Similarity 57.9%; Pred. No. 7.7e-34;
Matches 533; Conservative 0; Mismatches 387; Indels 0; Gaps 0;

```
Qy      22948 CCCCTCACCCTCCCCACCACCCACCAACCCCACCAAACCTGGCTCATCGCCATCCCCGAA 23007
              |||| | ||| |||| || ||| ||  |||| |   ||    | |  | ||  ||||
Db        926 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCACACCCCCCCCCCCCCACCCCCCC 867

Qy      23008 ACCCAGACCCACCACCCCCACATCACCAACATCCTCACCAACCTCCACCACCACGGCATC 23067
              |||   ||| || ||||| |  | ||  |  || | ||  || || ||  | |  | |
Db        866 CCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCC 807

Qy      23068 ACCCCCATCCCCCTCACTGTCAACCACACCCACACCAACCCCCAACACCTCCACCACACC 23127
              ||||  ||||| | |   | || | ||| | ||  |||||  | || || || | ||
Db        806 CNCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 747

Qy      23128 CTCCACCACACCCGACAACAAGCCCAAAACCACACCACCGGACCCATCACCGGCCTGCTC 23187
              | || || | |||  |  |  |||    || | || ||  ||||  | ||  ||  | |
Db        746 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCC 687

Qy      23188 TCCCTCCTCGCCCTCGACGAAACACCCCACCCCCACCACCCCCACACACCCACCGGCACC 23247
              ||| || | ||| |  |   | || ||||| || |||||| | |||  ||  | ||
Db        686 CCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCC 627

Qy      23248 CTCCTCAACCTCACCCTCCCCCAAACCCACACCCAAACCCACCCACCAACCCCCCTCTGG 23307
              |  | |  || | ||| |||||   ||| | |||   ||  ||| ||  ||| || |
Db        626 CCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCACCCCCCC 567

Qy      23308 TACGCCACCACCAACGCCACCACCACCCACCCCAACGACCCCCTCACACACCCCACCCAA 23367
              | || || ||  | || || ||||||  | ||  |  |||||  | | | |||| |||
Db        566 CCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCC 507

Qy      23368 GCCCAAACCTGGGGACTCGCCCGCACCACCCTCCTCGAACACCCCACCCACACCGCCGGA 23427
              |||   ||       | | ||| | || ||| || |   | |||| ||| | || ||
Db        506 CCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 447

Qy      23428 ATCATCGACCTCCCCACCACCCCCACCCCCCACACCCTCCACCACCTCACCCAAACCCTC 23487
              |  |  || |||| || |||||  |||||| | ||| || || || | |||   ||| |
Db        446 NNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 387

Qy      23488 ACCCAACCCCACCACCAAACCCAACTCGCCATCCGCACCACCGGCACCCACACCCGCCGC 23547
              |||  |||| || ||   |||  | | ||  || | || ||  | ||| | ||| || |
Db        386 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCC 327

Qy      23548 CTCACCCCCACCACCCTCACCCCCACACACCAACCACCCACCCCCACCCCCCACGGAACC 23607
              | | ||||| || ||| | ||||| | | ||   || ||| ||||| ||||||  |    ||
Db        326 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 267

Qy      23608 ACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCACC 23667
              |||  |  | ||  |   ||  | ||| | || ||| || | ||| || ||   |  ||
Db        266 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCTTCC 207

Qy      23668 CACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACGCA 23727
              | ||  ||| ||| || || || |  | ||  ||   ||  |||||  | |   || | |
Db        206 CCCCCCCCCCCCCCCACCCCCCCCCNCCCCCCNCCCCCCCCCCCCCCCCCCCGTTCCNCCCC 147

Qy      23728 CAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCGAC 23787
              |  || | |  | |||   |||  |      |  || || | ||  | | | |||  | |
Db        146 CNCCAGCCCNGCCCCCNGGTCCCCCCCNCGTCCCCCCCCCCCCCCCCCTCCCCCTTCCCC 87

Qy      23788 ACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCCTC 23847
              ||  |  |||   ||  | ||  |  | || |  | ||  ||||||  |  | ||||| |
Db         86 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCC 27

Qy      23848 ACCACCGTCATCCACACCGC 23867
              || ||  |  || | || |
Db         26 CCCCCCCCCCCCCCCCCCC 7
```

```
Qy       5269 CAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCGAC 5328
Db        425 CCCCCCCCCTCCCCCCCCCCCCCCCCCCTTCCTCCCCCCCCCCCCTCCCCNCCCCCCTCT 366

Qy       5329 ACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCCTC 5388
Db        365 CCCCNCCTNCCCCCCCCNCCTCCCCCCCTCCCCCCCCTCCCTCCCCCCCCCCCCCTCCCNC 306

Qy       5389 ACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCCCC 5448
Db        305 CCCCNCCNNGGGGGCTCCCCNNCCCCNCCCCCNNCNCCCACGGCCCCNCCCCCCCNCNCC 246

Qy       5449 ACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAACTC 5508
Db        245 CCCCCCCCCCCNNCCCCCNCCCCNTCCCCCCNCNCGGTCCNTCCCCCNCNCCNCCACGGC 186

Qy       5509 ACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCGGC 5568
Db        185 CNCNNGCTCNNCNNNCGGANCCCCCCCGCAACGTGGCCTCCCCCCCCCCACACCCGCGGC 126

Qy       5569 GCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACCAC 5628
Db        125 GCCCCCCCNCCCGCCCNCCCCTCCCNCCCCNNCCCCCCGCCCCCCCCCNCCCCCCCCCCC 66

Qy       5629 CGCCACACCCACCACCTCCCCGCCACCAGCATCGCC 5664
Db          65 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 30
```

RESULT 5
BF256757/c
LOCUS BF256757 1186 bp mRNA linear EST 22-OCT-2001
DEFINITION HVSMEf0010O19f Hordeum vulgare seedling root EST library HVcDNA0007 (Etiolated and unstressed) Hordeum vulgare cDNA clone HVSMEf0010O19f, mRNA sequence.
ACCESSION BF256757
VERSION BF256757.2 GI:16315529
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 1186)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi,D.W., Fenton,R.D., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex unstressed seedling root cDNA library
JOURNAL Unpublished (2001)
COMMENT On Nov 16, 2000 this sequence version replaced gi:11185870.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 216
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence start: 12
High quality sequence stop: 1164.
FEATURES Location/Qualifiers
source 1. .1186
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEf0010O19f"
/clone_lib="Hordeum vulgare seedling root EST library HVcDNA0007 (Etiolated and unstressed)"
/tissue_type="Seedling root"
/lab_host="TJC121"
/note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1; Seeds were surface sterilized then germinated under axenic conditions in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered crystallization dishes. Five-day old seedling roots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids. These steps were performed in the TJ Close laboratory at the University of California, Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT 47 a 32 c 948 g 118 t 41 others
ORIGIN

Query Match 1.1%; Score 323; DB 12; Length 1186;
Best Local Similarity 54.6%; Pred. No. 1.4e-35;
Matches 619; Conservative 0; Mismatches 514; Indels 1; Gaps 1;

```
Qy      18184 CACCGCCACCAACACGACCAAGCCCGCATCAACACCTGGACCTACCAGGAAACCTGGAAA 18243
Db       1181 CTCCCCCTCCTCCACCCCCCCCCCCTCCCCCCCCCCCCCACCCCTCCCCCCCCCCCCCAC 1122

Qy      18244 CCCCTCACCCTCCCCACCACCCACCAACCCCACCAAACCTGGCTCATCGCCATCCCCGAA 18303
Db       1121 CCCCTCCCCCCCCCCCCCCCTCACTCACCCCCCCCCCCCCCTCCCCCCCCCCCCCCACCCAC 1062

Qy      18304 ACCCAGACCCACCACCCCCACATCACCAACATCCTCACCAACCTCCACCACCACGGCATC 18363
Db       1061 AGCCCTCCACCCCACCCACCCCCCACCCCCACCTCCCCACCCACACCCCCCCCCCCACCCC 1002

Qy      18364 ACCCCCATCCCCCTCACCCTCAACCACACCCACACCAACCCCCAACACCTCCACCACACC 18423
Db       1001 TCCCACCCTCCCACCTCCCTACCACCCCACCCCCANCACCCCCCCCCCCCCCCCCCACCCC 942

Qy      18424 CGACAACAAGCCCAAAACCACACCACCGGACCCATCACCGGCCTGCTCTCCCTCCTCGCC 18483
Db        941 CCGCCCCCCCCCACCCCACCCCCCCCACACCCCCCCATCCCCCCCCCACCCCCACCCCACCCC 882

Qy      18484 CTCGACGAAACACCCCACCCCCACCACCCCCACACACCCACCGGCACCCTCCTCAACCTC 18543
Db        881 CCCACCCCAGCAACCTCACACCCCACCCCCCCCCCCCCCCACCACCACCCCCCCCCGCCAC 822

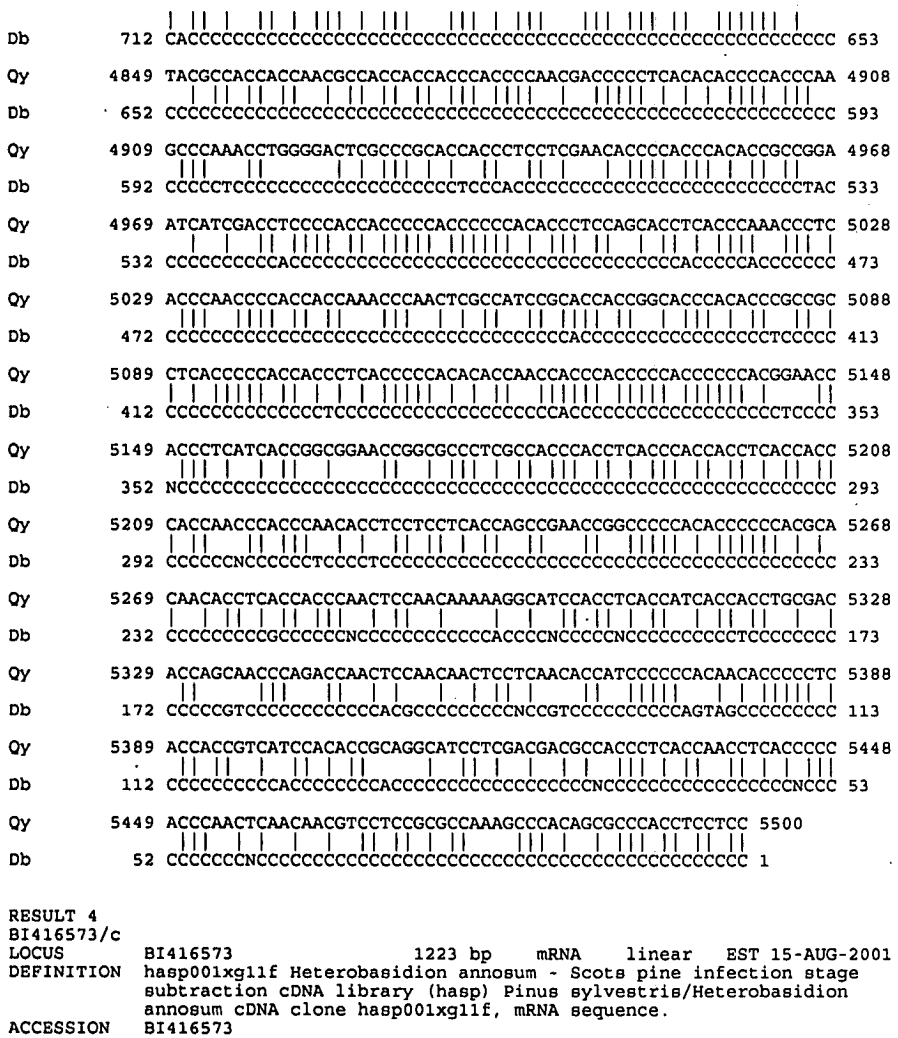
Qy      18544 ACCCTCACCCAAACCCACACCCAAACCCACCCACCAACCCCCCTCTGGTACGCCACCACC 18603
Db        821 CCCCCCCCCCCCACCCACCCCCCACCCCCCCACCCCCACCCCCCCCCCGAACCCCACCCCAC 762

Qy      18604 AACGCCACCACCACCCACCCCAACGACCCCCTCACACACCCCACCCAAGCCCAAACCTGG 18663
Db        761 ACCACACCCCACCCCCCCCCCACCCCGCCCCCCCCCCCACCCCCCCCCCCCACCACCCCCC 702

Qy      18664 GGACTCGCCCGCACCACCCTCCTCGAACACCCCACCCACACCGCCGGAATCATCGACCTC 18723
Db        701 CCACCCCCCAGCACCCCCCCCCCCCAACCCCCCACCCCCCACCACCCCCCCCCCCCAACCCC 642

Qy      18724 CCCACCACCCCCACCCCCCACACCCTCCACCACCTCACCCAAACCCTCACCCAACCCCAC 18783
Db        641 CCCCCCCCCCCCCCCCCACACCCCCCCCCCCCCCCCCACGACCCCCCCCCCACCCACCCACC 582

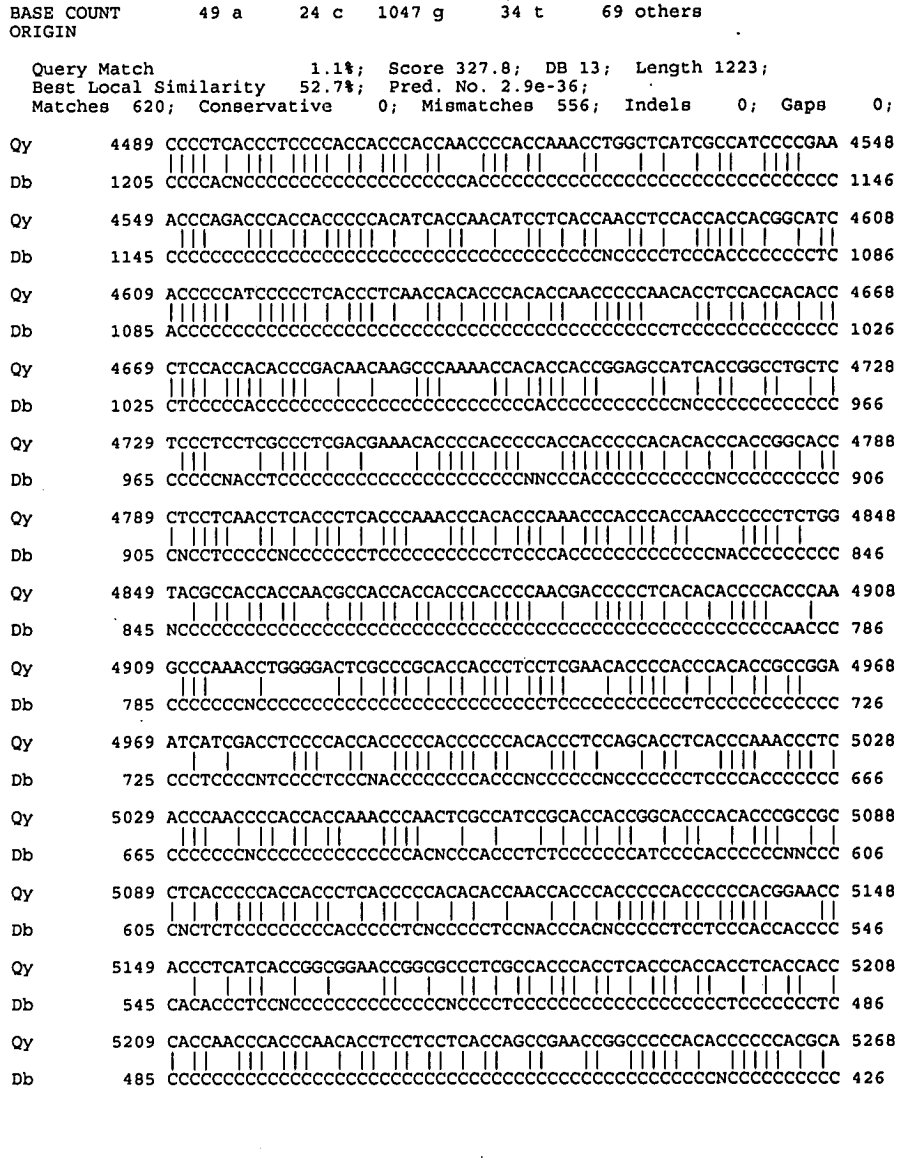
Qy      18784 CACCAAACCCAACTCGCCATCCGCACCACCGGCACCCACACCCGCCGCCTCACCCCCACC 18843
```

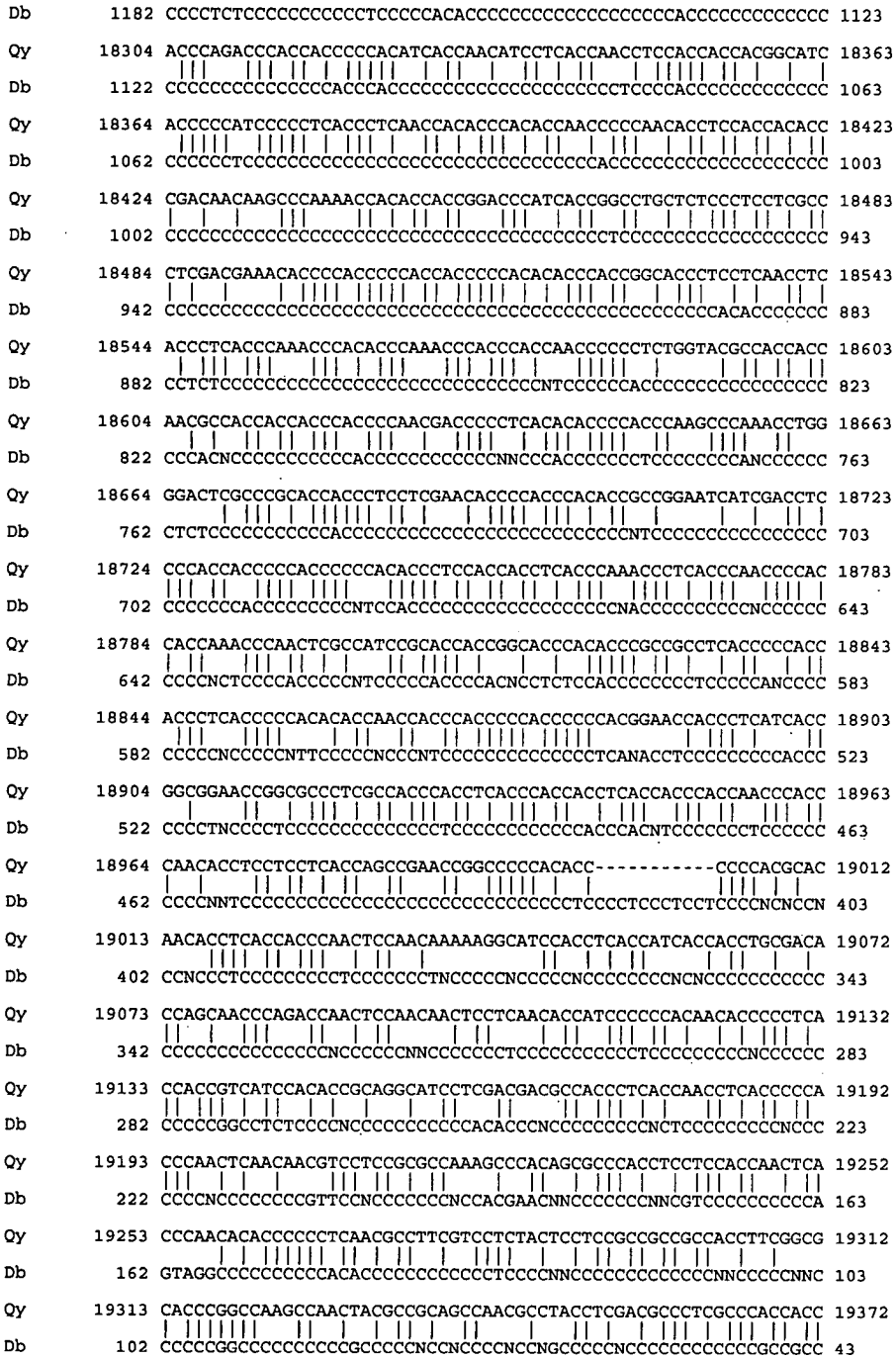


RESULT 4
BI416573/c
LOCUS BI416573 1223 bp mRNA linear EST 15-AUG-2001
DEFINITION hasp001xg11f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xg11f, mRNA sequence.
ACCESSION BI416573
VERSION BI416573.1 GI:15187596
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 1223)
AUTHORS Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026,S-750 07, Uppsala, Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se

Seq primer: T7 primer.
FEATURES Location/Qualifiers
source 1. .1223
/organism="Pinus sylvestris/Heterobasidion annosum"
/db_xref="taxon:169015"
/clone="hasp001xg11f"
/clone_lib="Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp)"
/dev_stage="Seedling roots of scots pine were infected for 6 days with H. annosum"
/note="Vector: pT-Adv; Site_1: EcoRI; The subtractive hybridization cDNA library was constructed from scots pine roots infected for 6-days with mycelia of Heterobasidon annosum (FP5)."
BASE COUNT 49 a 24 c 1047 g 34 t 69 others
ORIGIN

Query Match 1.1%; Score 327.8; DB 13; Length 1223;
Best Local Similarity 52.7%; Pred. No. 2.9e-36;
Matches 620; Conservative 0; Mismatches 556; Indels 0; Gaps 0;





```
Db      1182 CCCCTCTCCCCCCCCCCCCTCCCCCACACCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCC 1123

Qy     18304 ACCCAGACCCACCACCCCCACATCACCAACATCCTCACCAACCTCCACCACCACGGCATC 18363
             ||| |||  || | |||||  | ||  |  || | ||  | ||||| || |  |  |
Db      1122 CCCCCCCCCCCCCCCCACCCACCCCCCCCCCCCCCCCCCCCCCCTCCCCACCCCCCCCCCC 1063

Qy     18364 ACCCCCATCCCCCTCACCCTCAACCACACCCACACCAACCCCCAACACCTCCACCACACC 18423
             |||||  |||||  | ||| |  || | ||| | ||  | |||  | || || || | ||
Db      1062 CCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCC 1003

Qy     18424 CGACAACAAGCCCAAAACCACACCACCGGACCCATCACCGGCCTGCTCTCCCTCCTCGCC 18483
             |  |  |   |||    || | || ||   |||  | ||  ||  | | ||| || | ||
Db      1002 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCC 943

Qy     18484 CTCGACGAAACACCCCACCCCCACCACCCCCACACACCCACCGGCACCCTCCTCAACCTC 18543
             | |  |    | |||| ||||| || ||||| | | ||| ||  | |||  | |  || |
Db       942 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCACACCCCCCC 883

Qy     18544 ACCCTCACCCAAACCCACACCCAAACCCACCCACCAACCCCCCTCTGGTACGCCACCACC 18603
             | ||| |||   |||  | |||   ||| ||| |   |||||  |     | || || ||
Db       882 CCTCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNTCCCCCCACCCCCCCCCCCCCCCCC 823

Qy     18604 AACGCCACCACCACCCACCCCAACGACCCCCTCACACACCCCACCCAAGCCCAAACCTGG 18663
             | |  || || |||  |||  |  ||||  | ||| ||||  ||   ||||  ||
Db       822 CCCACNCCCCCCCCCCCACCCCCCCCCCCCNNCCCACCCCCCCTCCCCCCCCANCCCCCC 763

Qy     18664 GGACTCGCCCGCACCACCCTCCTCGAACACCCCACCCACACCGCCGGAATCATCGACCTC 18723
             |  ||| | |||||| || |   | |||| ||| | ||  |      |  | || |
Db       762 CTCTCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCNTCCCCCCCCCCCCCCCCC 703

Qy     18724 CCCACCACCCCCACCCCCCACACCCTCCACCACCTCACCCAAACCCTCACCCAACCCCAC 18783
             ||| ||  |||| |||| |||||  || || || | |||  |||| | |||  |||| |
Db       702 CCCCCCCACCCCCCCCCNTCCACCCCCCCCCCCCCCCCCCCCNACCCCCCCCCCCNCCCCCC 643

Qy     18784 CACCAAACCCAACTCGCCATCCGCACCACCGGCACCCACACCCGCCGCCTCACCCCCACC 18843
             | |||   ||| || | |   || |||| |   |  |  ||||| || |  | ||  | ||
Db       642 CCCCNCTCCCCACCCCCNTCCCCCACCCCACNCCTCTCCACCCCCCCCTCCCCCCANCCCC 583

Qy     18844 ACCCTCACCCCCACACACCAACCACCCACCCCCACCCCCCACGGAACCACCCTCATCACC 18903
             |||   ||||    | ||  ||  || ||||| |||||       | ||| |   ||
Db       582 CCCCCNCCCCCNTTCCCCCNCCCNTCCCCCCCCCCCCCCCTCANACCTCCCCCCCCCCACCC 523

Qy     18904 GGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCACCCACCAACCCACC 18963
             |    ||  | |||  | || ||| || | ||| ||  |  |||  ||| ||  ||| ||
Db       522 CCCCTNCCCCTCCCCCCCCCCCCCCTCCCCCCCCCCCCCACCCACNTCCCCCCCTCCCCCC 463

Qy     18964 CAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACC----------CCCCACGCAC 19012
             |  |    || || | ||  ||   ||  ||||| |  |          |||| | |
Db       462 CCCCNNTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCTCCCTCCTCCCCNCNCCN 403

Qy     19013 AACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCGACA 19072
             ||||  || |||  | ||  |          ||  | | ||     | ||  |  |
Db       402 CCNCCCTCCCCCCCCCTCCCCCCCTNCCCCCNCCCCCNCCCCCCCCCNCNCCCCCCCCCCC 343

Qy     19073 CCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCCTCA 19132
             ||  |  |||    ||  | ||       | ||    | ||  ||| || |  | | ||| |
Db       342 CCCCCCCCCCCCCCCNCCCCCCNNCCCCCCCTCCCCCCCCCCCTCCCCCCCCCNCCCCCC 283

Qy     19133 CCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCCCCA 19192
             || ||| | ||  | |  |   |  ||   ||    || ||| | |   || | || ||
Db       282 CCCCCGGCCTCTCCCCNCCCCCCCCCCCACACCCNCCCCCCCCCCNCTCCCCCCCCCNCCC 223

Qy     19193 CCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAACTCA 19252
             |||  | |   |   ||| || | ||   |  ||  | ||| |  | ||| ||  | ||
Db       222 CCCCNCCCCCCCCGTTCCNCCCCCCCCNCCACGAACNNCCCCCCCCNNCGTCCCCCCCCCCA 163

Qy     19253 CCCAACACACCCCCCTCAACGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCGGCG 19312
             | | |||||| || | ||  |  ||||  |  | || || || ||  |  |
Db       162 GTAGGCCCCCCCCCCACACCCCCCCCCCCCTCCCCNNCCCCCCCCCCCCCCNNCCCCCNNC 103

Qy     19313 CACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACCACC 19372
             | ||||||||   ||  |  | || |  ||      |  ||  |  ||| | ||| || ||
Db       102 CCCCCGGCCCCCCCCCCGCCCCCNCCNCCCCNCCNGCCCCCNCCCCCCCCCCCCCGCCGCC 43

Qy     19373 GCCACACCCACCACCTCCCCGCCACCAGCATCGCC 19407
             ||   ||| || || |||| |  ||  |  | ||
Db        42 CCCCNCCCCCCCCCCCCCCCCCNCCCCCCCCCCCC 8


RESULT 3
BI416650/c
LOCUS       BI416650                1015 bp    mRNA    linear   EST 15-AUG-2001
DEFINITION  hasp001xm01f Heterobasidion annosum - Scots pine infection stage
            subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
            annosum cDNA clone hasp001xm01f, mRNA sequence.
ACCESSION   BI416650
VERSION     BI416650.1  GI:15187673
KEYWORDS    EST.
SOURCE      Pinus sylvestris/Heterobasidion annosum.
  ORGANISM  Pinus sylvestris/Heterobasidion annosum
            Eukaryota; mixed EST libraries.
REFERENCE   1  (bases 1 to 1015)
  AUTHORS   Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
  TITLE     Expressed sequence tags of randomly selected cDNA clones from the
            interaction of the root rot fungus (Heterobasidion annosum) with
            seedling roots of Scots pine (Pinus sylvestris)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Fred O. Asiegbu
            Dept. of Forest Mycology & Pathology
            Swedish University of Agriculture, Box 7026,S-750 07, Uppsala,
            Sweden
            Tel: +46 18 67 15 98
            Fax: +46 18 30 92 45
            Email: Fred.Asiegbu@mykopat.slu.se
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .1015
                     /organism="Pinus sylvestris/Heterobasidion annosum"
                     /db_xref="taxon:169015"
                     /clone="hasp001xm01f"
                     /clone_lib="Heterobasidion annosum - Scots pine infection
                     stage subtraction cDNA library (hasp)"
                     /dev_stage="Seedling roots of scots pine were infected for
                     6 days with H. annosum"
                     /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive
                     hybridization cDNA library was constructed from  scots
                     pine roots infected for 6-days with mycelia of
                     Heterobasidon  annosum (FP5)."
BASE COUNT      21 a      6 c    961 g     16 t     11 others
ORIGIN

  Query Match             1.1%;  Score 329;  DB 13;  Length 1015;
  Best Local Similarity   57.4%;  Pred. No. 2e-36;
  Matches  581;  Conservative    0;  Mismatches  431;  Indels    0;  Gaps    0;

Qy      4489 CCCCTCACCCTCCCCACCACCCACCAACCCCACCAAACCTGGCTCATCGCCATCCCCGAA 4548
             |||||| ||| || | || ||| ||  |||| ||   |||    |  | ||  ||||
Db      1012 CCCCTCTCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCTCCTCCCCCCCCCCCCCCTC 953

Qy      4549 ACCCAGACCCACCACCCCCACATCACCAACATCCTCACCAACCTCCACCACCACGGCATC 4608
             |||   ||| || ||  | |  | ||| |  |||| ||  || || || || |  | |
Db       952 CCCCCCCCCCNCCCCCTACCCCCCCCCACCCCCCTCCCCCCCCCCCCCCCCCCCCCCCC 893

Qy      4609 ACCCCCATCCCCCTCACCCTCAACCACACCCACACCAACCCCCAACACCTCCACCACACC 4668
             | |||   |||| | ||| |  || | ||| | ||  |||||  | || || || | ||
Db       892 CCNCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 833

Qy      4669 CTCCACCACACCCGACAACAAGCCCAAAACCACACCACCGGAGCCATCACCGGCCTGCTC 4728
             | || || | |||  |  |   |||    || | || ||    ||  | ||  ||  | |
Db       832 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 773

Qy      4729 TCCCTCCTCGCCCTCGACGAAACACCCCACCCCCACCACCCCCACACACCCACCGGCACC 4788
             |||| || | ||| |  |     | |||| |||||| || |||||  | | ||| ||  | ||
Db       772 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 713

Qy      4789 CTCCTCAACCTCACCCTCACCCAAACCCACACCCAAACCCACCCACCAACCCCCCTCTGG 4848
```

```
                     /db_xref="taxon:169015"
                     /clone="hasp001xa01f"
                     /clone_lib="Heterobasidion annosum - Scots pine infection
                     stage subtraction cDNA library (hasp)"
                     /dev_stage="Seedling roots of scots pine were infected for
                     6 days with H. annosum"
                     /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive
                     hybridization cDNA library was constructed from  scots
                     pine roots infected for 6-days with mycelia of
                     Heterobasidon  annosum (FP5)."
BASE COUNT       20 a      28 c    1036 g      28 t      85 others
ORIGIN

  Query Match             1.1%;  Score 345.6;  DB 13;  Length 1197;
  Best Local Similarity   53.9%;  Pred. No. 9.5e-39;
  Matches  630;  Conservative    0;  Mismatches  537;  Indels    2;  Gaps    1;

Qy       4489 CCCCTCACCCTCCCCACCACCCACCAACCCCACCAAACCTGGCTCATCGCCATCCCCGAA 4548
              |||| | ||| | || ||  || ||  |||| ||   ||   | |  | ||  ||||
Db       1193 CCCCCCCCCCCCACCCCCCACCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCC 1134

Qy       4549 ACCCAGACCCACCACCCCCACATCACCAACATCCTCACCAACCTCCACCACCACGGCATC 4608
              |||   ||| || ||| | |  | ||  |  || | ||  || || || || | |  | |
Db       1133 CCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCNCCCCCNCCCC 1074

Qy       4609 ACCCCCATCCCCCTCACCCTCAACCACACCCACACCAACCCCCAACACCTCCACCACACC 4668
              |||||  |||||  | |||  |  |||  ||| | ||  |||||     || || || | ||
Db       1073 CCCCCCCCCCCCCNCNCCCCCCCCCAACCCCCCCCCCCCCCCCCNNCCCCCCCCCCCCCC 1014

Qy       4669 CTCCACCACACCCGACAACAAGCCCAAAACCACACCACCGGAGCCATCACCGGCCTGCTC 4728
              | || || | |   |  |   |||    || | || ||     |  | ||  ||  | |
Db       1013 CCCCCCCCCCCAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCNCCCCCCCCCCCC 954

Qy       4729 TCCCTCCTCGCCCTCGACGAAACACCCCACCCCCACCACCCCCACACACCCACCGGCACC 4788
              |||| || | ||| |  |       |||| |||||  | ||||| | |  ||| ||  | ||
Db        953 CCCCCCCCCCCCCCCCCCCCCNCACCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCC 894

Qy       4789 CTCCTCAACCTCACCCTCACCCAAACCCACACCCAAACCCACCCACCAACCCCCCTCTGG 4848
              | || |  || | ||  | |||   ||| | |||   ||| ||| ||  ||| || |
Db        893 CCCCCCCCCCCCCCCNCCCCCCCCCNCCCCCCCCCCCCNCCCCCCCCCCNCCCCNCCCCCN 834

Qy       4849 TACGCCACCACCAACGCCACCACCACCCACCCCAACGACCCCCTCACACACCCCACCCAA 4908
              | || || ||  | || || || ||| ||||  |  ||||| | | | |||| |||
Db        833 CCCCCCCCCCCCNNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCC 774

Qy       4909 GCCCAAACCTGGGGACTCGCCCGCACCACCCTCCTCGAACACCCCACCCACACCGCCGGA 4968
              |||   ||       | | | | | .|| ||| |       | ||||   | | || ||
Db        773 CCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCNNNCNCCCCCCCCCNNCCCCCCCCCCC 714

Qy       4969 ATCATCGACCTCCCCACCACCCCCACCCCCCACACCCTCCAGCACCTCACCCAAACCCTC 5028
                |  || |  | || |||||  ||||| | ||| ||  | || | |||    ||| |
Db        713 CNNNCCNCCCCCNNCCCCCCCCCCNNCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCC 654

Qy       5029 ACCCAACCCCACCACCAAACCCAACTCGCCATCCGCACCACCGGCACCCACACCCGCCGC 5088
              |||  |||| || ||   |||  | | ||  || | || ||  | ||| | ||| || |
Db        653 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCC 594

Qy       5089 CTCACCCCCACCACCCTCACCCCCACACACCAACCAC--CCACCCCCACCCCCCACGGAA 5146
              | | || || ||| |  || | | ||||| ||||  |||||||  || ||| |
Db        593 CCCCCCNCCCCCCCCCCCCNCCNCCCCCACCACCCACCACCCACCCCCCNCCNCCCCNNCC 534

Qy       5147 CCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCACCACCTCACCA 5206
              || ||| || | ||  |    ||  |  || | || ||| || | ||| || || | ||
Db        533 CCCCCCACACCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 474

Qy       5207 CCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACG 5266
              ||| ||  ||| |||  | || || || | ||  ||   ||  ||||| | |||||| |
Db        473 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNNCCCCCCCCCCCCCCCCCCCCCCCCN 414

Qy       5267 CACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCG 5326
              | |  | | | ||| |||  |  |  |       ||  || || | ||  | ||| ||  |
Db        413 CCCCCCNNCNCCCCCCCCCCCCGCCCCCCCCCCGCCCCCCCCCCCCCNCCCCCCCCCCCC 354
```

```
Qy       5327 ACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCCCCCACAACACCCCC 5386
              | |  |   ||   || |||| |  || | |  | ||  | ||| | |  | |
Db        353 CCCCTCCCCCTCCCTCCCCCCTCCCCCCCCCTTCCCCCCCCCTGCGCCGCGCTNCNTCAGG 294

Qy       5387 TCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCCACCCTCACCAACCTCACCC 5446
              | || |  |  |  || |  | || |  | |  |  || | ||  || | ||  |
Db        293 CCTCCCCNCNCCNNCCNCCCCCNNCAAACCCCNCCCCCCCCNCTCCCCNCCCNCCCCTNCC 234

Qy       5447 CCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAAC 5506
              || ||  | |    || ||  |||| ||||  | | |  |  | || ||  |
Db        233 CCTCCNCCCCCCGTCCNNCCCGCCNTNCCAAGGCCCCGCGTGGCGCCNAGCCCCCCCCGC 174

Qy       5507 TCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCG 5566
              |      | | ||||||   || ||  || ||     | |   | || ||  | ||  |
Db        173 CCGGGTGGCCCGCCCCCCCACCCCCCCCCGCCCATCCCCCNAGCCCCCCCCTCCCCCCCC 114

Qy       5567 GCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACC 5626
              |  |||| ||   ||  |  | ||    ||  |  |  || |  | ||| | | | |
Db        113 CCCNNCCCGCCCCCCCCCNCCCCCCCCCNCCCCCCCCNCCCCCCCCCCCCCCCCNCNCCCN 54

Qy       5627 ACCGCCACACCCACCACCTCCCCGCCACC 5655
              || || | ||| || || ||||  | ||
Db         53 NCCCCCCCCCCCCCCNCCCCCCCNNCCCC 25


RESULT 2
BI416537/c
LOCUS       BI416537                1224 bp    mRNA    linear   EST 15-AUG-2001
DEFINITION  hasp001xe15f Heterobasidion annosum - Scots pine infection stage
            subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion
            annosum cDNA clone hasp001xe15f, mRNA sequence.
ACCESSION   BI416537
VERSION     BI416537.1  GI:15187560
KEYWORDS    EST.
SOURCE      Pinus sylvestris/Heterobasidion annosum.
  ORGANISM  Pinus sylvestris/Heterobasidion annosum
            Eukaryota; mixed EST libraries.
REFERENCE   1  (bases 1 to 1224)
  AUTHORS   Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
  TITLE     Expressed sequence tags of randomly selected cDNA clones from the
            interaction of the root rot fungus (Heterobasidion annosum) with
            seedling roots of Scots pine (Pinus sylvestris)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Fred O. Asiegbu
            Dept. of Forest Mycology & Pathology
            Swedish University of Agriculture, Box 7026,S-750 07, Uppsala,
            Sweden
            Tel: +46 18 67 15 98
            Fax: +46 18 30 92 45
            Email: Fred.Asiegbu@mykopat.slu.se
            Seq primer: T7 primer.
FEATURES             Location/Qualifiers
     source          1. .1224
                     /organism="Pinus sylvestris/Heterobasidion annosum"
                     /db_xref="taxon:169015"
                     /clone="hasp001xe15f"
                     /clone_lib="Heterobasidion annosum - Scots pine infection
                     stage subtraction cDNA library (hasp)"
                     /dev_stage="Seedling roots of scots pine were infected for
                     6 days with H. annosum"
                     /note="Vector: pT-Adv; Site_1: EcoRI; The subtractive
                     hybridization cDNA library was constructed from  scots
                     pine roots infected for 6-days with mycelia of
                     Heterobasidon  annosum (FP5)."
BASE COUNT       50 a      14 c    1063 g      38 t      59 others
ORIGIN

  Query Match             1.1%;  Score 335.2;  DB 13;  Length 1224;
  Best Local Similarity   54.2%;  Pred. No. 2.7e-37;
  Matches  637;  Conservative    0;  Mismatches  527;  Indels   11;  Gaps    1;

Qy      18244 CCCCTCACCCTCCCCACCACCCACCAACCCCACCAAACCTGGCTCATCGCCATCCCCGAA 18303
              |||||| ||| |||| || ||| |  ||||| ||   ||   | |  | ||  ||||
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 01:08:39 ; Search time 24141 Seconds
(without alignments)
20589.018 Million cell updates/sec

Title: US-09-914-286-1
Perfect score: 30690
Sequence: 1 gtgcagaggatggacggcgg.....acgaaggtgagacagcgtga 30690

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

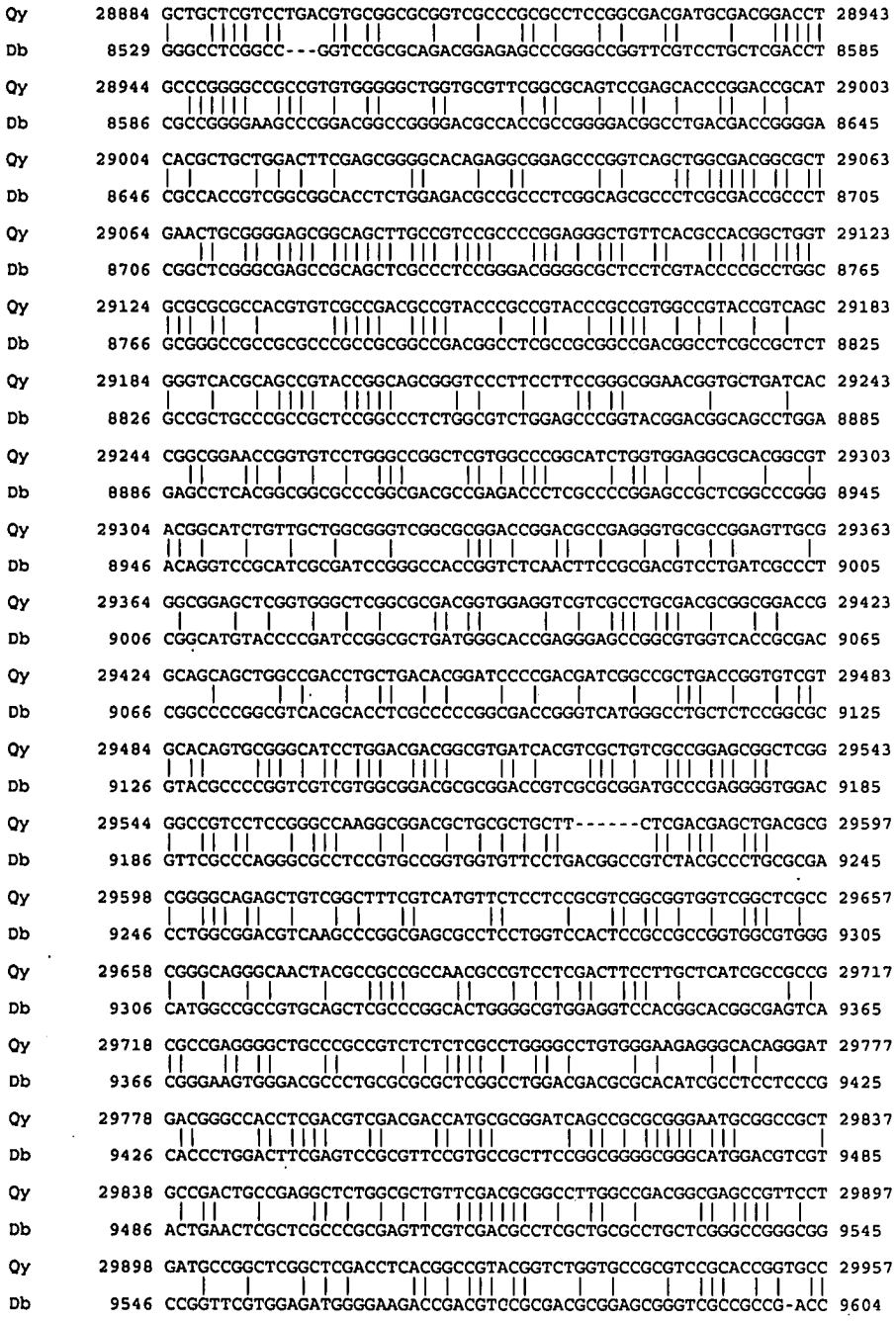
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

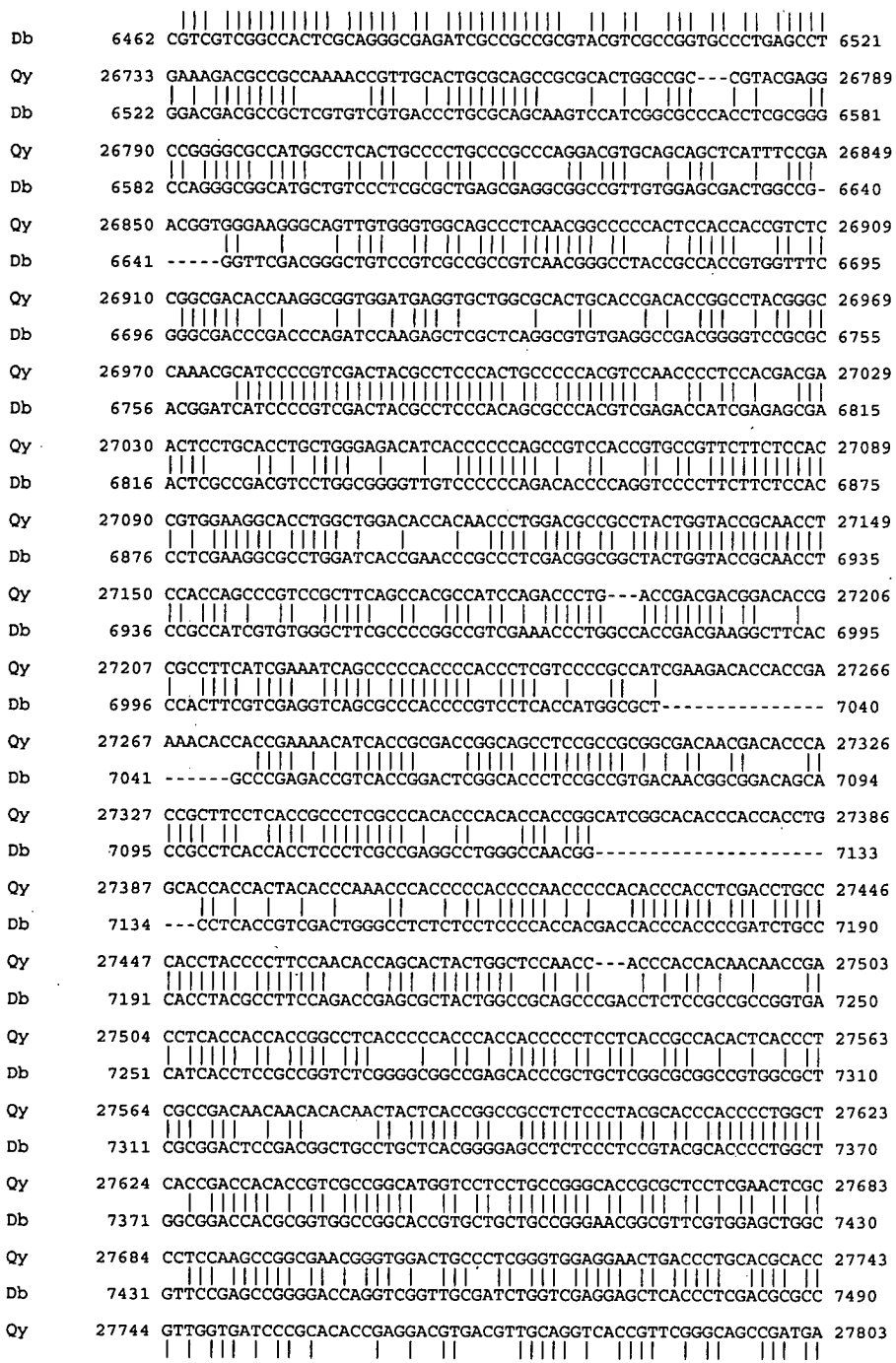
Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	345.6	1.1	1197	13	BI416470	BI416470 hasp001xa
c	2	335.2	1.1	1224	13	BI416537	BI416537 hasp001xe
c	3	329	1.1	1015	13	BI416650	BI416650 hasp001xm
c	4	327.8	1.1	1223	13	BI416573	BI416573 hasp001xg
c	5	323	1.1	1186	12	BF256757	BF256757 HVSMEf001
c	6	310.4	1.0	931	13	BI416575	BI416575 hasp001xg
c	7	306.4	1.0	1222	17	BH146899	BH146899 ENTQM47TF
c	8	301	1.0	1453	10	AW949571	AW949571 EST361641
c	9	299.6	1.0	1222	13	BI416660	BI416660 hasp001xm
c	10	299.2	1.0	1196	12	BF256755	BF256755 HVSMEf001
c	11	298.2	1.0	1042	13	BI416535	BI416535 hasp001xe
c	12	297.6	1.0	1082	10	BE455154	BE455154 HVSMEh009
c	13	297.6	1.0	1251	10	BE455089	BE455089 HVSMEh009
c	14	295.6	1.0	1166	12	BF256751	BF256751 HVSMEf001
c	15	295.2	1.0	1005	17	BH148610	BH148610 ENTPZ67TR
c	16	293.4	1.0	1138	17	AG152606	AG152606 Pan trogl
c	17	292	1.0	1321	17	AG126084	AG126084 Pan trogl
c	18	291	0.9	1138	12	BF256506	BF256506 HVSMEf001
c	19	290	0.9	1189	10	BE455091	BE455091 HVSMEh009
c	20	289.2	0.9	873	12	BG786249	BG786249 SEAUMC006
c	21	286.8	0.9	1277	12	BG447302	BG447302 GA__Eb004
	22	285.6	0.9	1606	13	BM470613	BM470613 AGENCOURT
	23	283.8	0.9	1161	14	BM912505	BM912505 AGENCOURT
c	24	282.2	0.9	1076	17	AG032994	AG032994 Pan trogl
c	25	282.2	0.9	1161	12	BF256588	BF256588 HVSMEf001
c	26	282.2	0.9	1195	13	BI416506	BI416506 hasp001xc
	27	282.4	0.9	1303	14	BM908878	BM908878 AGENCOURT
c	28	281.8	0.9	1152	12	BF256631	BF256631 HVSMEf001
c	29	280	0.9	1328	10	BE455227	BE455227 HVSMEh009
	30	279.2	0.9	1235	13	BM475329	BM475329 AGENCOURT
c	31	278.8	0.9	1177	17	AG081646	AG081646 Pan trogl
c	32	278.6	0.9	1294	12	BF256749	BF256749 HVSMEf001
c	33	277.6	0.9	1116	12	BF256617	BF256617 HVSMEf001
c	34	277.2	0.9	1184	14	BM911719	BM911719 AGENCOURT
c	35	277.4	0.9	1417	12	BF257444	BF257444 HVSMEf001
c	36	276.8	0.9	1297	14	BQ064843	BQ064843 AGENCOURT
c	37	275.8	0.9	1088	12	BF256580	BF256580 HVSMEf001
c	38	275.2	0.9	1137	17	AG078502	AG078502 Pan trogl
c	39	274.6	0.9	1293	12	BF256498	BF256498 HVSMEf001
c	40	273.4	0.9	877	13	BI416653	BI416653 hasp001xm
c	41	273.6	0.9	1292	12	BF255049	BF255049 HVSMEf000
c	42	273.2	0.9	1391	10	BE455155	BE455155 HVSMEh009
c	43	272.8	0.9	1036	12	BF254665	BF254665 HVSMEf000
c	44	271.6	0.9	1201	14	BQ063878	BQ063878 AGENCOURT
	45	270.2	0.9	960	14	BQ429353	BQ429353 AGENCOURT

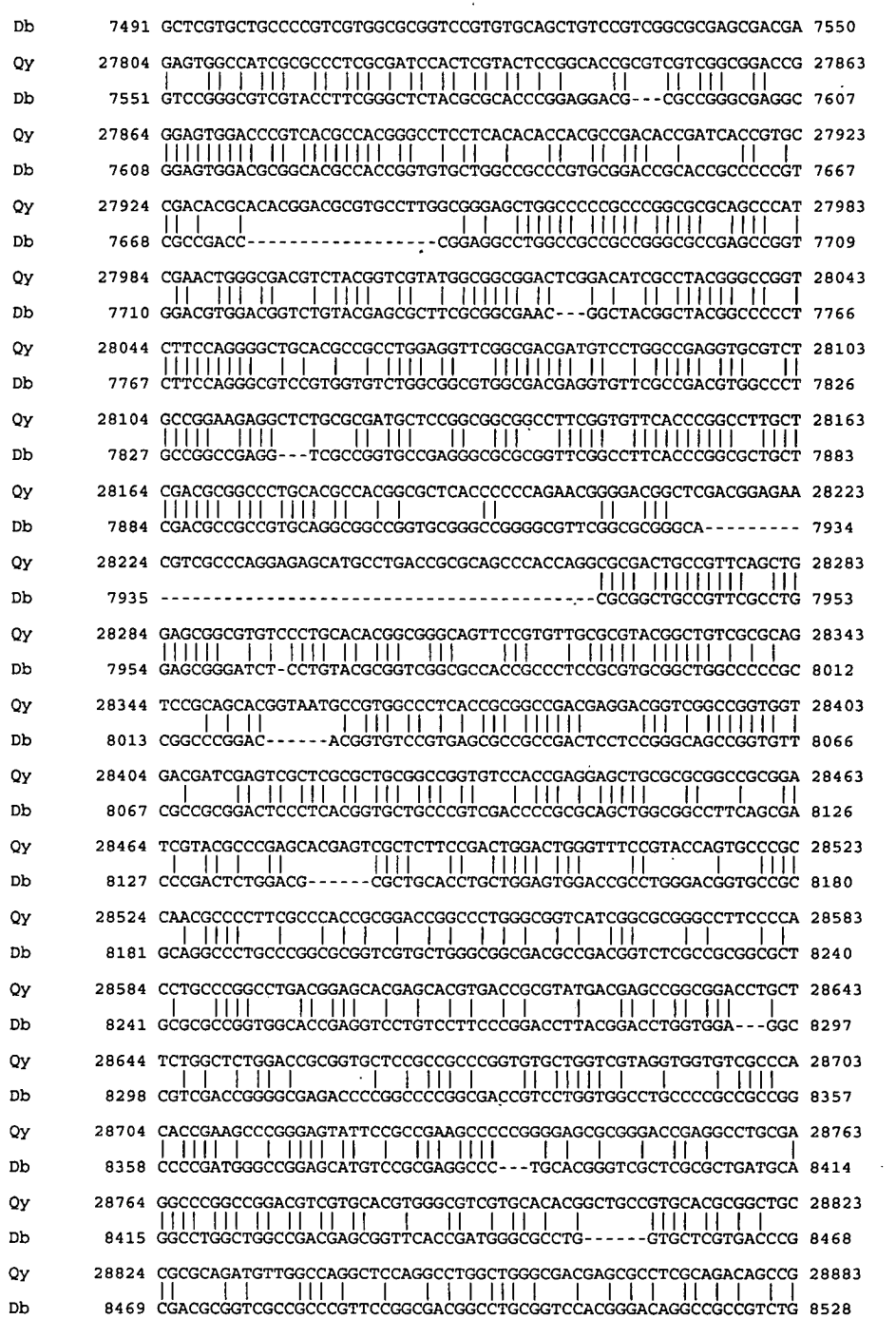
ALIGNMENTS

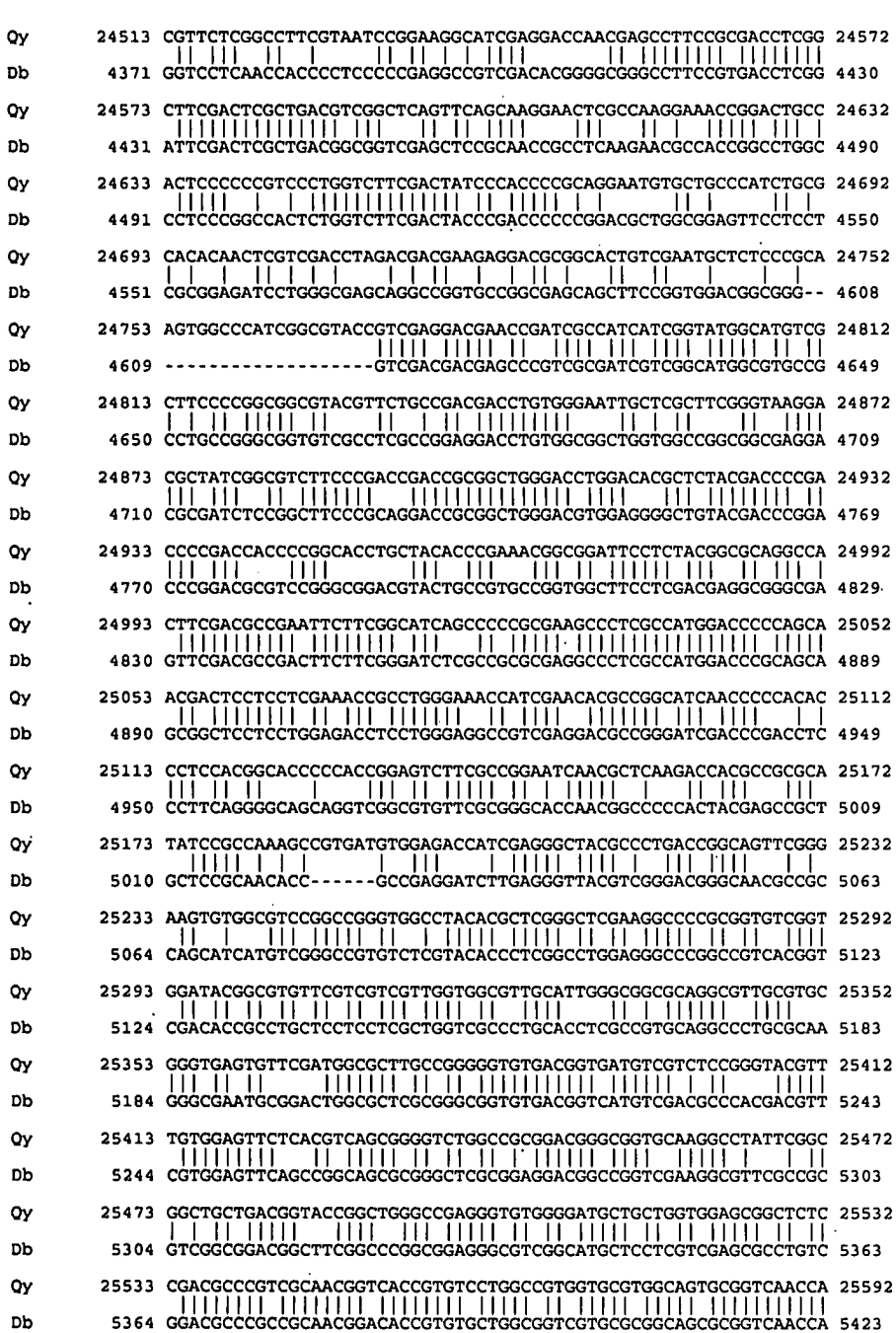
RESULT 1
BI416470/c
LOCUS BI416470 1197 bp mRNA linear EST 15-AUG-2001
DEFINITION hasp001xa01f Heterobasidion annosum - Scots pine infection stage subtraction cDNA library (hasp) Pinus sylvestris/Heterobasidion annosum cDNA clone hasp001xa01f, mRNA sequence.
ACCESSION BI416470
VERSION BI416470.1 GI:15187493
KEYWORDS EST.
SOURCE Pinus sylvestris/Heterobasidion annosum.
ORGANISM Pinus sylvestris/Heterobasidion annosum
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 1197)
AUTHORS Asiegbu,F.O., Nahalkova,J., Choi,W., Stenlid,J. and Dean,R.A.
TITLE Expressed sequence tags of randomly selected cDNA clones from the interaction of the root rot fungus (Heterobasidion annosum) with seedling roots of Scots pine (Pinus sylvestris)
JOURNAL Unpublished (2001)
COMMENT Contact: Fred O. Asiegbu
Dept. of Forest Mycology & Pathology
Swedish University of Agriculture, Box 7026,S-750 07, Uppsala, Sweden
Tel: +46 18 67 15 98
Fax: +46 18 30 92 45
Email: Fred.Asiegbu@mykopat.slu.se
Seq primer: T7 primer.
FEATURES Location/Qualifiers
source 1. .1197
/organism="Pinus sylvestris/Heterobasidion annosum"

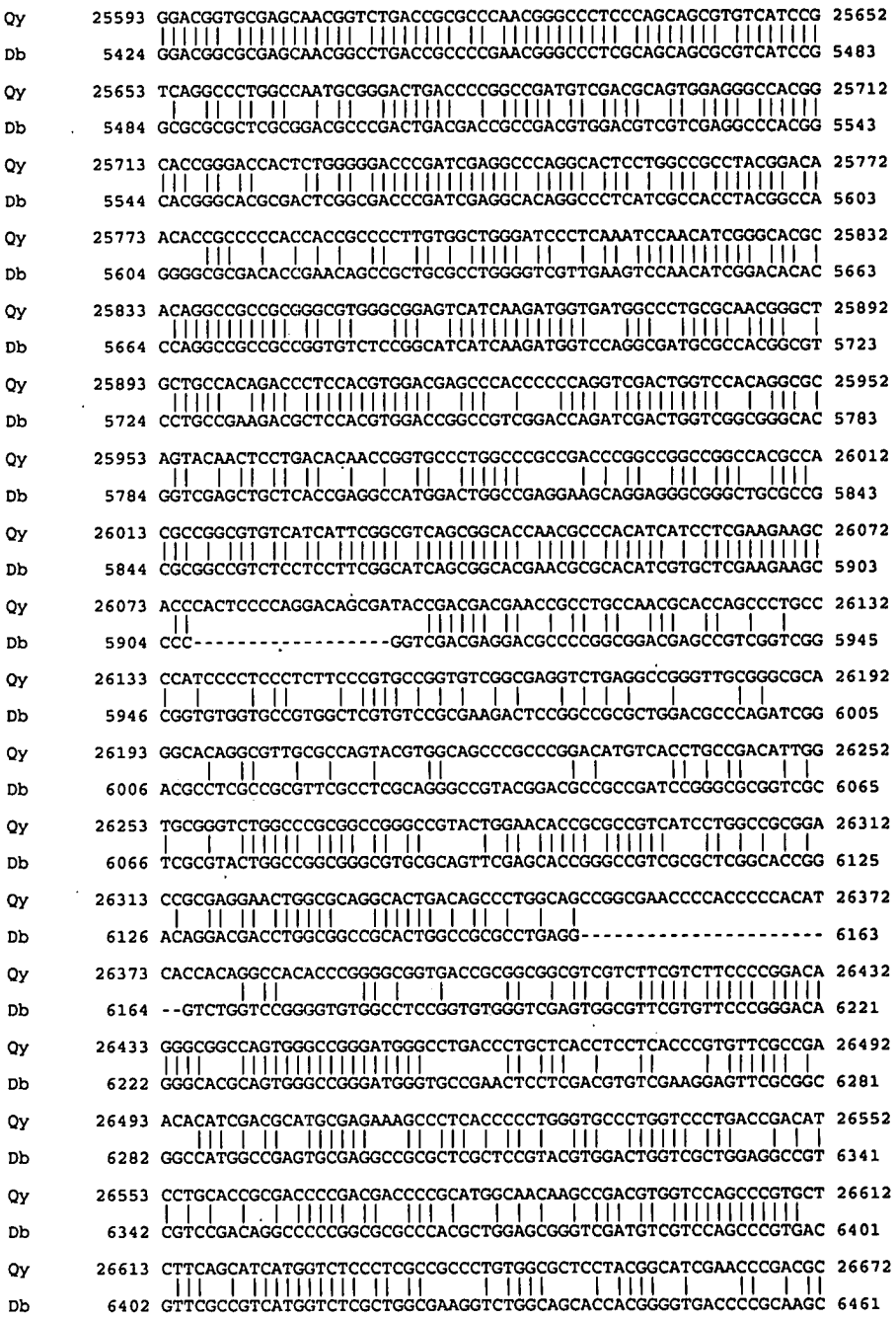


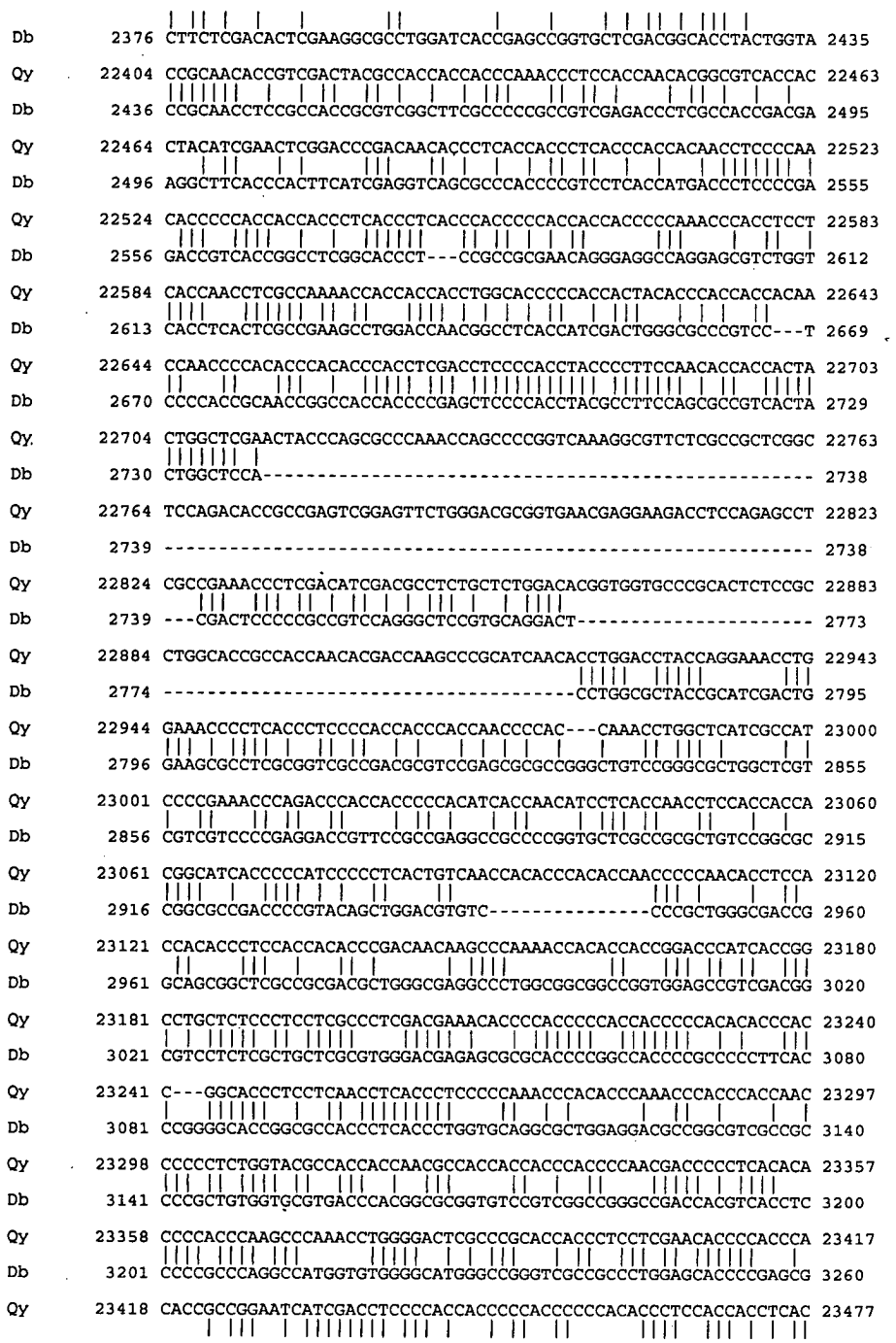
Search completed: June 19, 2003, 07:56:19
Job time : 4834 secs

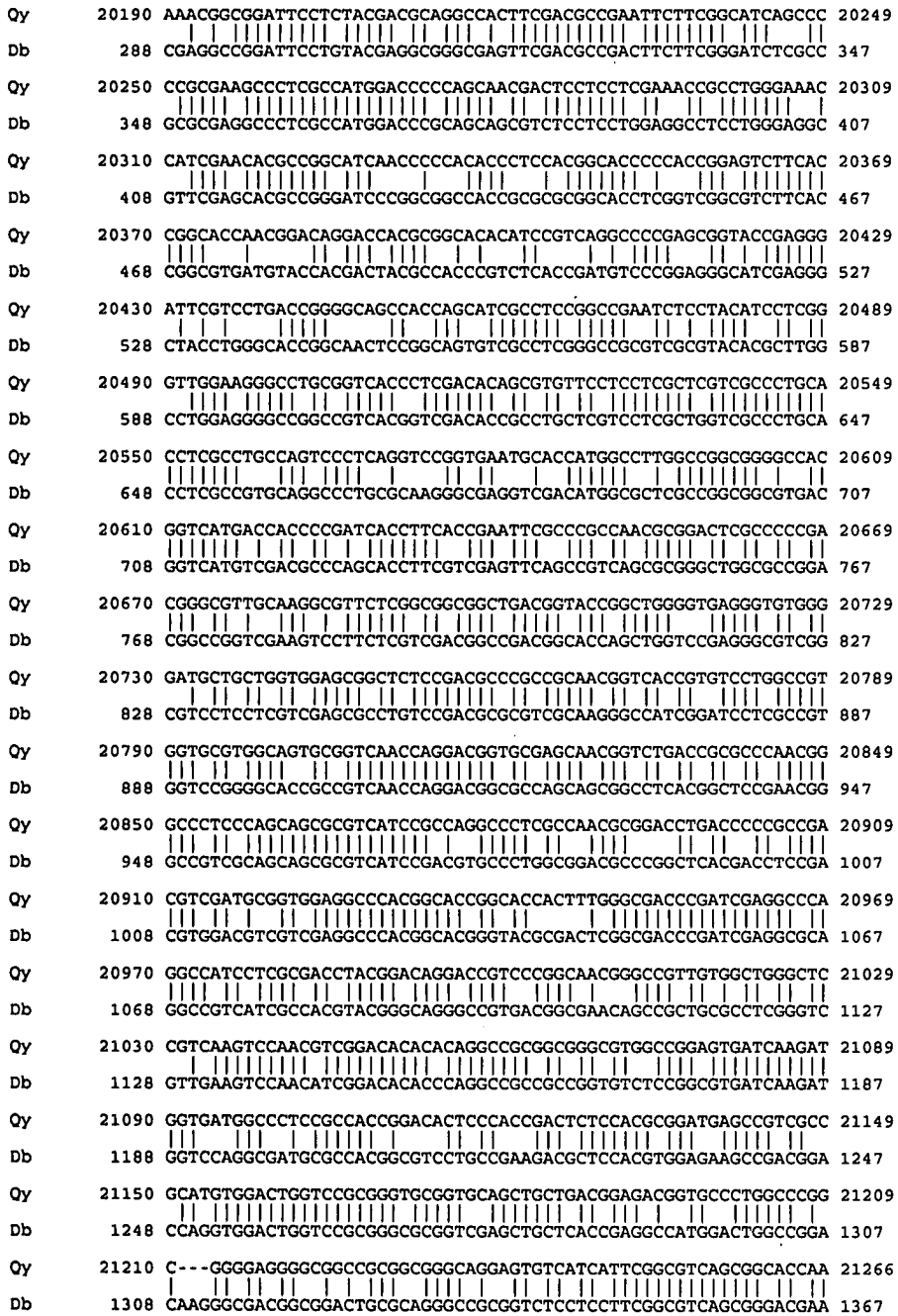

Db 6462 CGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTCGCCGGTGCCCTGAGCCT 6521
Qy 26733 GAAAGACGCCGCCAAAACCGTTGCACTGCGCAGCCGCGCACTGGCCGC---CGTACGAGG 26789
Db 6522 GGACGACGCCGCTCGTGTCGTGACCCTGCGCAGCAAGTCCATCGGCGCCCACCTCGCGGG 6581
Qy 26790 CCGGGGCGCCATGGCCTCACTGCCCCTGCCCGCCCAGGACGTGCAGCAGCTCATTTCCGA 26849
Db 6582 CCAGGGCGGCATGCTGTCCCTCGCGCTGAGCGAGGCGGCCGTTGTGGAGCGACTGGCCG- 6640
Qy 26850 ACGGTGGGAAGGGCAGTTGTGGGTGGCAGCCCTCAACGGCCCCCACTCCACCACCGTCTC 26909
Db 6641 -----GGTTCGACGGGCTGTCCGTCGCCGCCGTCAACGGGCCTACCGCCACCGTGGTTTC 6695
Qy 26910 CGGCGACACCAAGGCGGTGGATGAGGTGCTGGCGCACTGCACCGACACCGGCCTACGGGC 26969
Db 6696 GGGCGACCCGACCCAGATCCAAGAGCTCGCTCAGGCGTGTGAGGCCGACGGGGTCCGCGC 6755
Qy 26970 CAAACGCATCCCCGTCGACTACGCCTCCCACTGCCCCCACGTCCAACCCCTCCACGACGA 27029
Db 6756 ACGGATCATCCCCGTCGACTACGCCTCCCACAGCGCCCACGTCGAGACCATCGAGAGCGA 6815
Qy 27030 ACTCCTGCACCTGCTGGGAGACATCACCCCCCAGCCGTCCACCGTGCCGTTCTTCTCCAC 27089
Db 6816 ACTCGCCGACGTCCTGGCGGGGTTGTCCCCCCAGACACCCCAGGTCCCCTTCTTCTCCAC 6875
Qy 27090 CGTGGAAGGCACCTGGCTGGACACCACAACCCTGGACGCCGCCTACTGGTACCGCAACCT 27149
Db 6876 CCTCGAAGGCGCCTGGATCACCGAACCCGCCCTCGACGGCGGCTACTGGTACCGCAACCT 6935
Qy 27150 CCACCAGCCCGTCCGCTTCAGCCACGCCATCCAGACCCTG---ACCGACGACGGACACCG 27206
Db 6936 CCGCCATCGTGTGGGCTTCGCCCCGGCCGTCGAAACCCTGGCCACCGACGAAGGCTTCAC 6995
Qy 27207 CGCCTTCATCGAAATCAGCCCCCACCCCACCCTCGTCCCCGCCATCGAAGACACCACCGA 27266
Db 6996 CCACTTCGTCGAGGTCAGCGCCCACCCCGTCCTCACCATGGCGCT--------------- 7040
Qy 27267 AAACACCACCGAAAACATCACCGCGACCGGCAGCCTCCGCCGCGGCGACAACGACACCCA 27326
Db 7041 ------GCCCGAGACCGTCACCGGACTCGGCACCCTCCGCCGTGACAACGGCGGACAGCA 7094
Qy 27327 CCGCTTCCTCACCGCCCTCGCCCACACCCACACCACCGGCATCGGCACACCCACCACCTG 27386
Db 7095 CCGCCTCACCACCTCCCTCGCCGAGGCCTGGGCCAACGG---------------------- 7133
Qy 27387 GCACCACCACTACACCCAAACCCACCCCCACCCCAACCCCCACACCCACCTCGACCTGCC 27446
Db 7134 ---CCTCACCGTCGACTGGGCCTCTCTCCTCCCCACCACGACCACCCACCCCGATCTGCC 7190
Qy 27447 CACCTACCCCTTCCAACACCAGCACTACTGGCTCCAACC---ACCCACCACAACAACCGA 27503
Db 7191 CACCTACGCCTTCCAGACCGAGCGCTACTGGCCGCAGCCCGACCTCTCCGCCGCCGGTGA 7250
Qy 27504 CCTCACCACCACCGGCCTCACCCCCACCCACCACCCCCTCCTCACCGCCACACTCACCCT 27563
Db 7251 CATCACCTCCGCCGGTCTCGGGGCGGCCGAGCACCCGCTGCTCGGCGCGGCCGTGGCGCT 7310
Qy 27564 CGCCGACAACAACACACAACTACTCACCGGCCGCCTCTCCCTACGCACCCACCCCTGGCT 27623
Db 7311 CGCGGACTCCGACGGCTGCCTGCTCACGGGGAGCCTCTCCCTCCGTACGCACCCCTGGCT 7370
Qy 27624 CACCGACCACACCGTCGCCGGCATGGTCCTCCTGCCGGGCACCGCGCTCCTCGAACTCGC 27683
Db 7371 GGCGGACCACGCGGTGGCCGGCACCGTGCTGCTGCCGGGAACGGCGTTCGTGGAGCTGGC 7430
Qy 27684 CCTCCAAGCCGGCGAACGGGTGGACTGCCCTCGGGTGGAGGAACTGACCCTGCACGCACC 27743
Db 7431 GTTCCGAGCCGGGGACCAGGTCGGTTGCGATCTGGTCGAGGAGCTCACCCTCGACGCGCC 7490
Qy 27744 GTTGGTGATCCCGCACACCGAGGACGTGACGTTGCAGGTCACCGTTCGGGCAGCCGATGA 27803

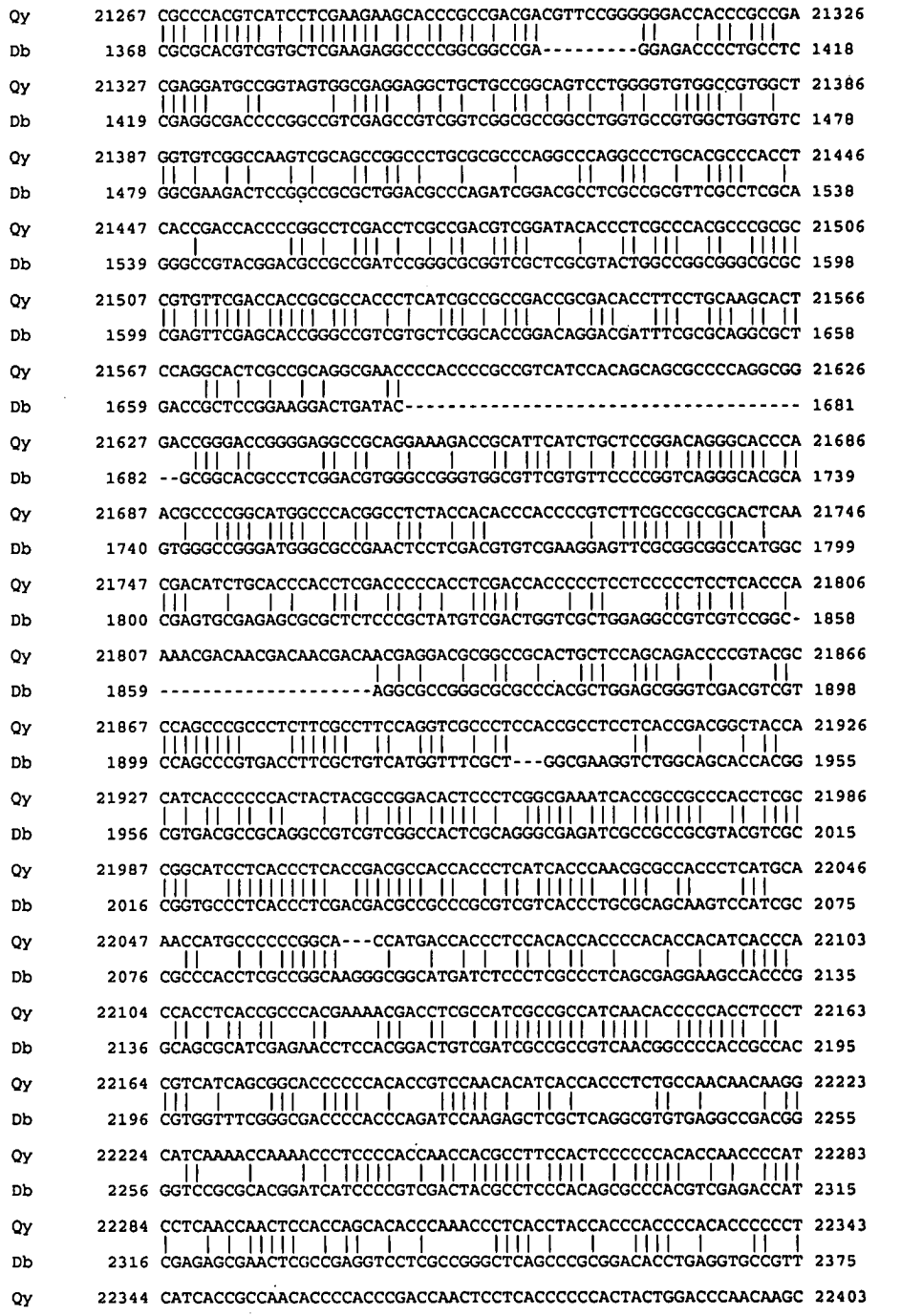

Db 7491 GCTCGTGCTGCCCCGTCGTGGCGCGGTCCGTGTGCAGCTGTCCGTCGGCGCGAGCGACGA 7550
Qy 27804 GAGTGGCCATCGCGCCCTCGCGATCCACTCGTACTCCGGCACCGCGTCGTCGGCGGACCG 27863
Db 7551 GTCCGGGCGTCGTACCTTCGGGCTCTACGCGCACCCGGAGGACG---CGCCGGGCGAGGC 7607
Qy 27864 GGAGTGGACCCGTCACGCCACGGGCCTCCTCACACACCACGCCGACACCGATCACCGTGC 27923
Db 7608 GGAGTGGACGCGGCACGCCACCGGTGTGCTGGCCGCCCGTGCGGACCGCACCGCCCCCGT 7667
Qy 27924 CGACACGCACACGGACGCGTGCCTTGGCGGGAGCTGGCCCCCGCCCGGCGCGCAGCCCAT 27983
Db 7668 CGCCGACC----------CGGAGGCCTGGCCGCCGCCGGGCGCCGAGCCGGT 7709
Qy 27984 CGAACTGGGCGACGTCTACGGTCGTATGGCGGCGGACTCGGACATCGCCTACGGGCCGGT 28043
Db 7710 GGACGTGGACGGTCTGTACGAGCGCTTCGCGGCGAAC---GGCTACGGCTACGGCCCCCT 7766
Qy 28044 CTTCCAGGGGCTGCACGCCGCCTGGAGGTTCGGCGACGATGTCCTGGCCGAGGTGCGTCT 28103
Db 7767 CTTCCAGGGCGTCCGTGGTGTCTGGCGGCGTGGCGACGAGGTGTTCGCCGACGTGGCCCT 7826
Qy 28104 GCCGGAAGAGGCTCTGCGCGATGCTCCGGCGGCGGCCTTCGGTGTTCACCCGGCCTTGCT 28163
Db 7827 GCCGGCCGAGG---TCGCCGGTGCCGAGGGCGCGCGGTTCGGCCTTCACCCGGCGCTGCT 7883
Qy 28164 CGACGCGGCCCTGCACGCCACGGCGCTCACCCCCCAGAACGGGGACGGCTCGACGGAGAA 28223
Db 7884 CGACGCCGCCGTGCAGGCGGCCGGTGCGGGCCGGGGCGTTCGGCGCGGGCA--------- 7934
Qy 28224 CGTCGCCCAGGAGAGCATGCCTGACCGCGCAGCCCACCAGGCGCGACTGCCGTTCAGCTG 28283
Db 7935 --CGCGGCTGCCGTTCGCCTG 7953
Qy 28284 GAGCGGCGTGTCCCTGCACACGGCGGGCAGTTCCGTGTTGCGCGTACGGCTGTCGCGCAG 28343
Db 7954 GAGCGGGATCT-CCTGTACGCGGTCGGCGCCACCGCCCTCCGCGTGCGGCTGGCCCCCGC 8012
Qy 28344 TCCGCAGCACGGTAATGCCGTGGCCCTCACCGCGGCCGACGAGGACGGTCGGCCGGTGGT 28403
Db 8013 CGGCCCGGAC------ACGGTGTCCGTGAGCGCCGCCGACTCCTCCGGGCAGCCGGTGTT 8066
Qy 28404 GACGATCGAGTCGCTCGCGCTGCGGCCGGTGTCCACCGAGGAGCTGCGCGCGGCCGCGGA 28463
Db 8067 CGCCGCGGACTCCCTCACGGTGCTGCCCGTCGACCCCGCGCAGCTGGCGGCCTTCAGCGA 8126
Qy 28464 TCGTACGCCCGAGCACGAGTCGCTCTTCCGACTGGACTGGGTTTCCGTACCAGTGCCCGC 28523
Db 8127 CCCGACTCTGGACG------CGCTGCACCTGCTGGAGTGGACCGCCTGGGACGGTGCCGC 8180
Qy 28524 CAACGCCCCTTCGCCCACCGCGGACCGGCCCTGGGCGGTCATCGGCGCGGGCCTTCCCCA 28583
Db 8181 GCAGGCCCTGCCCGGCGCGGTCGTGCTGGGCGGCGACGCCGACGGTCTCGCCGCGGCGCT 8240
Qy 28584 CCTGCCCGGCCTGACGGAGCACGAGCACGTGACCGCGTATGACGAGCCGGCGGACCTGCT 28643
Db 8241 GCGCGCCGGTGGCACCGAGGTCCTGTCCTTCCCGGACCTTACGGACCTGGTGGA---GGC 8297
Qy 28644 TCTGGCTCTGGACCGCGGTGCTCCGCCGCCCGGTGTGCTGGTCGTAGGTGGTGTCGCCCA 28703
Db 8298 CGTCGACCGGGGCGAGACCCCGGCCCCGGCGACCGTCCTGGTGGCCTGCCCCGCCGCCGG 8357
Qy 28704 CACCGAAGCCCGGGAGTATTCCGCCGAAGCCCCCGGGGAGCGCGGGACCGAGGCCTGCGA 28763
Db 8358 CCCCGATGGGCCGGAGCATGTCCGCGAGGCCC---TGCACGGGTCGCTCGCGCTGATGCA 8414
Qy 28764 GGCCCGGCCGGACGTCGTGCACGTGGGCGTCGTGCACACGGCTGCCGTGCACGCGGCTGC 28823
Db 8415 GGCCTGGCTGGCCGACGAGCGGTTCACCGATGGGCGCCTG------GTGCTCGTGACCCG 8468
Qy 28824 CGCGCAGATGTTGGCCAGGCTCCAGGCCTGGCTGGGCGACGAGCGCCTCGCAGACAGCCG 28883
Db 8469 CGACGCGGTCGCCGCCCGTTCCGGCGACGGCCTGCGGTCCACGGGACAGGCCGCCGTCTG 8528


Qy 24513 CGTTCTCGGCCTTCGTAATCCGGAAGGCATCGAGGACCAACGAGCCTTCCGCGACCTCGG 24572
Db 4371 GGTCCTCAACCACCCCTCCCCCGAGGCCGTCGACACGGGGCGGGCCTTCCGTGACCTCGG 4430
Qy 24573 CTTCGACTCGCTGACGTCGGCTCAGTTCAGCAAGGAACTCGCCAAGGAAACCGGACTGCC 24632
Db 4431 ATTCGACTCGCTGACGGCGGTCGAGCTCCGCAACCGCCTCAAGAACGCCACCGGCCTGGC 4490
Qy 24633 ACTCCCCCCGTCCCTGGTCTTCGACTATCCCACCCCGCAGGAATGTGCTGCCCATCTGCG 24692
Db 4491 CCTCCCGGCCACTCTGGTCTTCGACTACCCGACCCCCCGGACGCTGGCGGAGTTCCTCCT 4550
Qy 24693 CACACAACTCGTCGACCTAGACGACGAAGAGGACGCGGCACTGTCGAATGCTCTCCCGCA 24752
Db 4551 CGCGGAGATCCTGGGCGAGCAGGCCGGTGCCGGCGAGCAGCTTCCGGTGGACGGCGGG-- 4608
Qy 24753 AGTGGCCCATCGGCGTACCGTCGAGGACGAACCGATCGCCATCATCGGTATGGCATGTCG 24812
Db 4609 ------------------GTCGACGACGAGCCCGTCGCGATCGTCGGCATGGCGTGCCG 4649
Qy 24813 CTTCCCCGGCGGCGTACGTTCTGCCGACGACCTGTGGGAATTGCTCGCTTCGGGTAAGGA 24872
Db 4650 CCTGCCGGGCGGTGTCGCCTCGCCGGAGGACCTGTGGCGGCTGGTGGCCGGCGGCGAGGA 4709
Qy 24873 CGCTATCGGCGTCTTCCCGACCGACCGCGGCTGGGACCTGGACACGCTCTACGACCCCGA 24932
Db 4710 CGCGATCTCCGGCTTCCCGCAGGACCGCGGCTGGGACGTGGAGGGGCTGTACGACCCGGA 4769
Qy 24933 CCCCGACCACCCCGGCACCTGCTACACCCGAAACGGCGGATTCCTCTACGGCGCAGGCCA 24992
Db 4770 CCCGGACGCGTCCGGGCGGACGTACTGCCGTGCCGGTGGCTTCCTCGACGAGGCGGGCGA 4829
Qy 24993 CTTCGACGCCGAATTCTTCGGCATCAGCCCCCGCGAAGCCCTCGCCATGGACCCCCAGCA 25052
Db 4830 GTTCGACGCCGACTTCTTCGGGATCTCGCCGCGCGAGGCCCTCGCCATGGACCCGCAGCA 4889
Qy 25053 ACGACTCCTCCTCGAAACCGCCTGGGAAACCATCGAACACGCCGGCATCAACCCCCACAC 25112
Db 4890 GCGGCTCCTCCTGGAGACCTCCTGGGAGGCCGTCGAGGACGCCGGGATCGACCCGACCTC 4949
Qy 25113 CCTCCACGGCACCCCCACCGGAGTCTTCGCCGGAATCAACGCTCAAGACCACGCCGCGCA 25172
Db 4950 CCTTCAGGGGCAGCAGGTCGGCGTGTTCGCGGGCACCAACGGCCCCCACTACGAGCCGCT 5009
Qy 25173 TATCCGCCAAAGCCGTGATGTGGAGACCATCGAGGGCTACGCCCTGACCGGCAGTTCGGG 25232
Db 5010 GCTCCGCAACACC------GCCGAGGATCTTGAGGGTTACGTCGGGACGGGCAACGCCGC 5063
Qy 25233 AAGTGTGGCGTCCGGCCGGGTGGCCTACACGCTCGGGCTCGAAGGCCCCGCGGTGTCGGT 25292
Db 5064 CAGCATCATGTCGGGCCGTGTCTCGTACACCCTCGGCCTGGAGGGCCCGGCCGTCACGGT 5123
Qy 25293 GGATACGGCGTGTTCGTCGTCGTTGGTGGCGTTGCATTGGGCGGCGCAGGCGTTGCGTGC 25352
Db 5124 CGACACCGCCTGCTCCTCCTCGCTGGTCGCCCTGCACCTCGCCGTGCAGGCCCTGCGCAA 5183
Qy 25353 GGGTGAGTGTTCGATGGCGCTTGCCGGGGGTGTGACGGTGATGTCGTCTCCGGGTACGTT 25412
Db 5184 GGGCGAATGCGGACTGGCGCTCGCGGGCGGTGTGACGGTCATGTCGACGCCCACGACGTT 5243
Qy 25413 TGTGGAGTTCTCACGTCAGCGGGGTCTGGCCGCGGACGGGCGGTGCAAGGCCTATTCGGC 25472
Db 5244 CGTGGAGTTCAGCCGGCAGCGCGGGCTCGCGGAGGACGGCCGGTCGAAGGCGTTCGCCGC 5303
Qy 25473 GGCTGCTGACGGTACCGGCTGGGCCGAGGGTGTGGGGATGCTGCTGGTGGAGCGGCTCTC 25532
Db 5304 GTCGGCGGACGGCTTCGGCCCGGCGGAGGGCGTCGGCATGCTCCTCGTCGAGCGCCTGTC 5363
Qy 25533 CGACGCCCGTCGCAACGGTCACCGTGTCCTGGCCGTGGTGCGTGGCAGTGCGGTCAACCA 25592
Db 5364 GGACGCCCGCCGCAACGGACACCGTGTGCTGGCGGTCGTGCGCGGCAGCGCGGTCAACCA 5423


Qy 25593 GGACGGTGCGAGCAACGGTCTGACCGCGCCCAACGGGCCCTCCCAGCAGCGTGTCATCCG 25652
Db 5424 GGACGGCGCGAGCAACGGCCTGACCGCCCCGAACGGGCCCTCGCAGCAGCGCGTCATCCG 5483
Qy 25653 TCAGGCCCTGGCCAATGCGGGACTGACCCCGGCCGATGTCGACGCAGTGGAGGGCCACGG 25712
Db 5484 GCGCGCGCTCGCGGACGCCCGACTGACGACCGCCGACGTGGACGTCGTCGAGGCCCACGG 5543
Qy 25713 CACCGGGACCACTCTGGGGGACCCGATCGAGGCCCAGGCACTCCTGGCCGCCTACGGACA 25772
Db 5544 CACGGGCACGCGACTCGGCGACCCGATCGAGGCACAGGCCCTCATCGCCACCTACGGCCA 5603
Qy 25773 ACACCGCCCCCACCACCGCCCCTTGTGGCTGGGATCCCTCAAATCCAACATCGGGCACGC 25832
Db 5604 GGGGCGCGACACCGAACAGCCGCTGCGCCTGGGGTCGTTGAAGTCCAACATCGGACACAC 5663
Qy 25833 ACAGGCCGCCGCGGGCGTGGGCGGAGTCATCAAGATGGTGATGGCCCTGCGCAACGGGCT 25892
Db 5664 CCAGGCCGCCGCCGGTGTCTCCGGCATCATCAAGATGGTCCAGGCGATGCGCCACGGCGT 5723
Qy 25893 GCTGCCACAGACCCTCCACGTGGACGAGCCCACCCCCCAGGTCGACTGGTCCACAGGCGC 25952
Db 5724 CCTGCCGAAGACGCTCCACGTGGACCGGCCGTCGGACCAGATCGACTGGTCGGCGGGCAC 5783
Qy 25953 AGTACAACTCCTGACACAACCGGTGCCCTGGCCCGCCGACCCGGCCGGCCGGCCACGCCA 26012
Db 5784 GGTCGAGCTGCTCACCGAGGCCATGGACTGGCCGAGGAAGCAGGAGGGCGGGCTGCGCCG 5843
Qy 26013 CGCCGGCGTGTCATCATTCGGCGTCAGCGGCACCAACGCCCACATCATCCTCGAAGAAGC 26072
Db 5844 CGCGGCCGTCTCCTCCTTCGGCATCAGCGGCACGAACGCGCACATCGTGCTCGAAGAAGC 5903
Qy 26073 ACCCACTCCCCAGGACAGCGATACCGACGACGAACCGCCTGCCAACGCACCAGCCCTGCC 26132
Db 5904 CCC--------------------GGTCGACGAGGACGCCCCGGCGGACGAGCCGTCGGTCGG 5945
Qy 26133 CCATCCCCTCCCTCTTCCCGTGCCGGTGTCGGCGAGGTCTGAGGCCGGGTTGCGGGCGCA 26192
Db 5946 CGGTGTGGTGCCGTGGCTCGTGTCCGCGAAGACTCCGGCCGCGCTGGACGCCCAGATCGG 6005
Qy 26193 GGCACAGGCGTTGCGCCAGTACGTGGCAGCCCGCCCGGACATGTCACCTGCCGACATTGG 26252
Db 6006 ACGCCTCGCCGCGTTCGCCTCGCAGGGCCGTACGGACGCCGCCGATCCGGGCGCGGTCGC 6065
Qy 26253 TGCGGGTCTGGCCCGCGGCCGGGCCGTACTGGAACACCGCGCCGTCATCCTGGCCGCGGA 26312
Db 6066 TCGCGTACTGGCCGGCGGGCGTGCGCAGTTCGAGCACCGGGCCGTCGCGCTCGGCACCGG 6125
Qy 26313 CCGCGAGGAACTGGCGCAGGCACTGACAGCCCTGGCAGCCGGCGAACCCCACCCCCACAT 26372
Db 6126 ACAGGACGACCTGGCGGCCGCACTGGCCGCGCCTGAGG---------------------- 6163
Qy 26373 CACCACAGGCCACACCCGGGGCGGTGACCGCGGCGGCGTCGTCTTCGTCTTCCCCGGACA 26432
Db 6164 --GTCTGGTCCGGGGTGTGGCCTCCGGTGTGGGTCGAGTGGCGTTCGTGTTCCCGGGACA 6221
Qy 26433 GGGCGGCCAGTGGGCCGGGATGGGCCTGACCCTGCTCACCTCCTCACCCGTGTTCGCCGA 26492
Db 6222 GGGCACGCAGTGGGCCGGGATGGGTGCCGAACTCCTCGACGTGTCGAAGGAGTTCGCGGC 6281
Qy 26493 ACACATCGACGCATGCGAGAAAGCCCTCACCCCCTGGGTGCCCTGGTCCCTGACCGACAT 26552
Db 6282 GGCCATGGCCGAGTGCGAGGCCGCGCTCGCTCCGTACGTGGACTGGTCGCTGGAGGCCGT 6341
Qy 26553 CCTGCACCGCGACCCCGACGACCCCGCATGGCAACAAGCCGACGTGGTCCAGCCCGTGCT 26612
Db 6342 CGTCCGACAGGCCCCCGGCGCGCCCACGCTGGAGCGGGTCGATGTCGTCCAGCCCGTGAC 6401
Qy 26613 CTTCAGCATCATGGTCTCCCTCGCCGCCCTGTGGCGCTCCTACGGCATCGAACCCGACGC 26672
Db 6402 GTTCGCCGTCATGGTCTCGCTGGCGAAGGTCTGGCAGCACCACGGGGTGACCCCGCAAGC 6461
Qy 26673 GGTCCTCGGCCACTCCCAGGGAGAAATCGCCGCCGCCCACATCTGCGGCGCACTCAGCCT 26732


Db 2376 CTTCTCGACACTCGAAGGCGCCTGGATCACCGAGCCGGTGCTCGACGGCACCTACTGGTA 2435
Qy 22404 CCGCAACACCGTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCAC 22463
Db 2436 CCGCAACCTCCGCCACCGCGTCGGCTTCGCCCCCGCCGTCGAGACCCTCGCCACCGACGA 2495
Qy 22464 CTACATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAA 22523
Db 2496 AGGCTTCACCCACTTCATCGAGGTCAGCGCCCACCCCGTCCTCACCATGACCCTCCCCGA 2555
Qy 22524 CACCCCCACCACCACCCTCACCCTCACCCACCCCCACCACCACCCCCAAACCCACCTCCT 22583
Db 2556 GACCGTCACCGGCCTCGGCACCCT---CCGCCGCGAACAGGGAGGCCAGGAGCGTCTGGT 2612
Qy 22584 CACCAACCTCGCCAAAACCACCACCACCTGGCACCCCCACCACTACACCCACCACCACAA 22643
Db 2613 CACCTCACTCGCCGAAGCCTGGACCAACGGCCTCACCATCGACTGGGCGCCCGTCC---T 2669
Qy 22644 CCAACCCCACACCCACACCCACCTCGACCTCCCCACCTACCCCTTCCAACACCACCACTA 22703
Db 2670 CCCCACCGCAACCGGCCACCACCCCGAGCTCCCCACCTACGCCTTCCAGCGCCGTCACTA 2729
Qy 22704 CTGGCTCGAACTACCCAGCGCCCAAACCAGCCCCGGTCAAAGGCGTTCTCGCCGCTCGGC 22763
Db 2730 CTGGCTCCA---------- 2738
Qy 22764 TCCAGACACCGCCGAGTCGGAGTTCTGGGACGCGGTGAACGAGGAAGACCTCCAGAGCCT 22823
Db 2739 ---------- 2738
Qy 22824 CGCCGAAACCCTCGACATCGACGCCTCTGCTCTGGACACGGTGGTGCCCGCACTCTCCGC 22883
Db 2739 ---CGACTCCCCCGCCGTCCAGGGCTCCGTGCAGGACT---------- 2773
Qy 22884 CTGGCACCGCCACCAACACGACCAAGCCCGCATCAACACCTGGACCTACCAGGAAACCTG 22943
Db 2774 ----------CCTGGCGCTACCGCATCGACTG 2795
Qy 22944 GAAACCCCTCACCCTCCCCACCACCCACCAACCCCAC---CAAACCTGGCTCATCGCCAT 23000
Db 2796 GAAGCGCCTCGCGGTCGCCGACGCGTCCGAGCGCGCCGGGCTGTCCGGGCGCTGGCTCGT 2855
Qy 23001 CCCCGAAACCCAGACCCACCACCCCCACATCACCAACATCCTCACCAACCTCCACCACCA 23060
Db 2856 CGTCGTCCCCGAGGACCGTTCCGCCGAGGCCGCCCCGGTGCTCGCCGCGCTGTCCGGCGC 2915
Qy 23061 CGGCATCACCCCCATCCCCCTCACTGTCAACCACACCCACACCAACCCCCAACACCTCCA 23120
Db 2916 CGGCGCCGACCCCGTACAGCTGGACGTGTC----------CCCGCTGGGCGACCG 2960
Qy 23121 CCACACCCTCCACCACACCCGACAACAAGCCCAAAACCACACCACCGGACCCATCACCGG 23180
Db 2961 GCAGCGGCTCGCCGCGACGCTGGGCGAGGCCCTGGCGGCGGCCGGTGGAGCCGTCGACGG 3020
Qy 23181 CCTGCTCTCCCTCCTCGCCCTCGACGAAACACCCCACCCCCACCACCCCCACACACCCAC 23240
Db 3021 CGTCCTCTCGCTGCTCGCGTGGGACGAGAGCGCGCACCCCGGCCACCCCGCCCCCTTCAC 3080
Qy 23241 C---GGCACCCTCCTCAACCTCACCCTCCCCCAAACCCACACCCAAACCCACCCACCAAC 23297
Db 3081 CCGGGGCACCGGCGCCACCCTCACCCTGGTGCAGGCGCTGGAGGACGCCGGCGTCGCCGC 3140
Qy 23298 CCCCCTCTGGTACGCCACCACCAACGCCACCACCACCCACCCCAACGACCCCCTCACACA 23357
Db 3141 CCCGCTGTGGTGCGTGACCCACGGCGCGGTGTCCGTCGGCCGGGCCGACCACGTCACCTC 3200
Qy 23358 CCCCACCCAAGCCCAAACCTGGGGACTCGCCCGCACCACCCTCCTCGAACACCCCACCCA 23417
Db 3201 CCCCGCCCAGGCCATGGTGTGGGGCATGGGCCGGGTCGCCGCCCTGGAGCACCCCGAGCG 3260
Qy 23418 CACCGCCGGAATCATCGACCTCCCCACCACCCCCACCCCCCACACCCTCCACCACCTCAC 23477
Db 3261 GTGGGGCGGCCTGATCGACCTGCCCTCGGACGCCGACCGGGCGGCCCTGGACCGCATGAC 3320
Qy 23478 CCAAACCCTCACCCAACCCCACCACCAAACCCAACTCGCCATCCGCACCACCGGCACCCA 23537
Db 3321 CACGGTCCTCGCCGGCGGTACGGGTGAGGACCAGGTCGCGGTACGCGCCTCCGGGCTGCT 3380
Qy 23538 CACCCGCCGCCTCACCCCCACCACCCTCACCCCCACACACCAACCACCCACCCCCACCC- 23596
Db 3381 CGCCCGCCGCCTCGTCCGCGCCTCCCTCCCGGCGCACGGCACGGCTTCGCCGTGGTGGCA 3440
Qy 23597 --CCCACGGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACCCA 23654
Db 3441 GGCCGACGGCACGGTGCTCGTCACCGGTGCCGAGGAGCCTGCGGCCGCCGAGGCCGCACG 3500
Qy 23655 CCACCTCACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCCCA 23714
Db 3501 CCGGCTGGCCCGCGACGGCGCCGGACACCTCCTCCTCCACACCACCCCCTCCGGCAGCGA 3560
Qy 23715 CACCCCCCACGCACAACAC----------CTCACCACCCA 23744
Db 3561 AGGCGCCGAAGGCACCTCCGGTGCCGCCGAGGACTCCGGCCTCGCCGGGCTCGTCGCCGA 3620
Qy 23745 ACTCCAACAAAAAGGCATCCACCTCACCATCACCACCTGCGACACCAGCAACCCAGACCA 23804
Db 3621 ACTCGCGGACCTGGGCGCGACGGCCACCGTCGTGACCTGCGACCTCACGGACGCGGAGGC 3680
Qy 23805 ACTCCAACAACTCCTCAACACCATCCCCCCACAACACCCCCTCACCACCGTCATCCACAC 23864
Db 3681 GGCCGCCCGGCTGCTCGCCGGCGTCTCCGACGCGCACCCGCTCAGCGCCGTCCTCCACCT 3740
Qy 23865 CGCAGGCGTCAATCTCTTCGCCCCCGTGTCGGAAACCGATGCCGAATCCTTCTCTTCCGT 23924
Db 3741 GCCGCCCACCGTCGACTCCGAGCCGCTCGCCGCGACCGACGCGGACGCGCTCGCCCGTGT 3800
Qy 23925 TACGGCAGCGAAGGCAACGGGCGCGGCGATTCTGCATGAGTTGCTGCTGGACCATGAAAC 23984
Db 3801 CGTGACCGCGAAGGCCACCGCCGCGCTCCACCTGGACCGCCTCCTGCGGGAGGCCGCGGC 3860
Qy 23985 GCTTGAA----------CACTTCATTCTCTTCTCGTCGGGCGCCGGCGCTTGGGGCAG 24032
Db 3861 TGCCGGAGGCCGTCCGCCCGTCCTGGTCCTCTTCTCCTCGGTCGCCGCGATCTGGGGCGG 3920
Qy 24033 CGGGAATCAGTGCGCATACTCGGCGGCCAACGCATACCTGGACGCGCTCGCGACGCATCG 24092
Db 3921 CGCCGGTCAGGGCGCGTACGCCGCCGGTACGGCCTTCCTCGACGCCCTCGCCGGTCAGCA 3980
Qy 24093 TCAGACACATGGACTTCCCGGGGCATCGATCGCCTGGGGCCCCTGGGCCGGAAAGGGCAT 24152
Db 3981 CCGGGCCGACGGCCCCACCGTGACCTCGGTGGCCTGGAGCCCCTGGGAGGGCAGCCGCGT 4040
Qy 24153 GTCGGCCGGTGATGCGGCTCATGGTTACCTGGAAAAGCGCGGCATTCTGCCGATGGAGCC 24212
Db 4041 CACCGAGGGTGCGACCGGGGAGCGG---CTGCGCCGCCTCGGCCTGCGCCCCCTCGCCCC 4097
Qy 24213 ACGCATGGCGCTCGCGGCATTCCATCGTGCGCGGGCGCAGCGGCCGAATTCCAACCTGAT 24272
Db 4098 CGCGACGGCGCTCACCGCCCTGGACACCGCGCTCGGCCACGGCGACA---CCGCCGTCAC 4154
Qy 24273 CATCGCGGACATCGACTGGGAGCGCTTCGTCCCCGCCTTCACCGCTCGACGCCACAGCCC 24332
Db 4155 GATCGCCGACGTCGACTGGTCGAGCTTCGCCCCCGGCTTCACCACGGCCCGGCCGGGCAC 4214
Qy 24333 GCTCATCGAGGACATTCCGGAGGTTCGGCAAGCGGCTCAGGAGCTGGAAGCAGCTGCGTC 24392
Db 4215 CCTCCTCGCCGATCTGCCCGAGGCGCGCCGCGCGCTCGACGAGCAGCAGTCGACGACGGC 4274
Qy 24393 GACGGCAAAGACGACCACAGCTCAGCCGATTGCGACGTCTCTCCGTGAGCGATTGGCCCG 24452
Db 4275 CGCCGACGACACCGTCCTGAGCCGCGAGCTCGGTGCG---------- 4311
Qy 24453 ACTGACGTCCTCAAAGCAGAACCAGGTGCTGCTCGGCCTGATTCGGACAGGCATCTGCAC 24512
Db 4312 -CTCACCGGCGCCGAACAGCAGCGCCGTATGCAGGAGTTGGTCCGCGAGCACCTCGCCGT 4370





```
Db      11534 GTACCCGCCGGGGTCCTGGCGACGCTGTCCCTGATGCAGGCTATGGAGGAGGAGGCGGTG 11593

Qy      23293 CCAACCCCCCTCTGGTACGCCACCACCAACGCCACCACCACCCACCCCAACGACCCCCTC 23352

Db      11594 GAGGCTCGCGTGTGGTGCGTCTCCCGCGCCGCGGTCGCCGCCGCCGACCGGGAACGGCCC 11653

Qy      23353 ACACACCCCACCCAAGCCCAAACCTGGGGACTCGCCCGCACCACCCTCCTCGAACACCCC 23412

Db      11654 GTCGGCGCGGGCGCCGCCC---TGTGGGGGCTGGGGCGGGTGGCCGCCCTGGAACGCCCC 11710

Qy      23413 ACCCACACCGCCGGAATCATCGACCTCCCCACCACCCCCACCCCCCACACCCTCCACCAC 23472

Db      11711 ACCCGGTGGGGCGGTCTCGTGGACCTGCCCGCCTCGCCCGGTGCGGCGCACTGGGCGGCC 11770

Qy      23473 CTCACCCAAACCCTCACCCAACCCCACCACCAAACCCAACTCGCCATCCGCACCACCGGC 23532

Db      11771 GCCGTGGAACGGCTCGCCGGTCCCGAGGACCAGATCGCCGTGCGCGCGTCCGGCAGTTGG 11830

Qy      23533 ACCCACACCCGCCGCCTCACCCCCACCACCCTCACCCCCACACACCAACCACCCACCCCC 23592

Db      11831 GGCCGGCGCCTCACCAGGCTGCCGCGCGACGGCGGCGGCCGGACGGCCGCACCCGCGTAC 11890

Qy      23593 ACCCCCCACGGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACC 23652

Db      11891 CGGCCGCGCGGCACGGTGCTCGTCACCGGTGGCACCGGCGCGCTCGGCGGGCATCTCGCC 11950

Qy      23653 CACCACCTCACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCC 23712

Db      11951 CGCTGGCTCGCCGC---GGCGGGCGCCGAACACCTGGCGCTCACCAGCCGCCGGGGCCCG 12007

Qy      23713 CACACCCCCCACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACC 23772

Db      12008 GACGCGCCCGGCGCCGCCGGACTCGAGGCCGAACTCCTCCTCCTGGGCGCCAAGGTGACG 12067

Qy      23773 ATCACCACCTGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCC 23832

Db      12068 TTCGCCGCCTGCGACACCGCCGACCGCGACGGCCTCGCCCGGGTCCTGCGGGCGATACCG 12127

Qy      23833 CCACAACACCCCCTCACCACCGTCATCCACACCGCAGGCGTCAATCTCTTCGCCCCCGTG 23892

Db      12128 GAGGACACCCCGCTCACCGCGGTGTTCCACGCCGCGGGCGTACCGCAGGTCACGCCGCTG 12187

Qy      23893 TCGGAAACCGATGCCGAATCCTTCTCTTCCGTTACGGCAGCGAAGGCAACGGGCGCGGCG 23952

Db      12188 TCCCGTACCTCGCCCGAGCACTTCGCCGACGTGTACGCGGGCAAGGCGGCGGGCGCCGCG 12247

Qy      23953 ATTCTGCATGAGTTGCTGCTGGACCATGAAAC---GCTTGAACACTTCATTCTCTTCTCG 24009

Db      12248 CACCTGGACGAACTGACCCGCGAACTCGGCGCCGGACTCGACGCGTTCGTCCTCTACTCC 12307

Qy      24010 TCGGGCGCCGGCGCTTGGGGCAGCGGGAATCAGTGCGCATACTCGGCGGCCAACGCATAC 24069

Db      12308 TCCGGCGCCGGCGTCTGGGGCAGCGCCGGCCAGGGTGCCTACGCCGCCGCCAACGCCGCC 12367

Qy      24070 CTGGACGCGCTCGCGACGCATCGTCAGACACATGGACTTCCCGGGGCATCGATCGCCTGG 24129

Db      12368 CTGGACGCGCTCGCCCGGCGCCGTGCGGCGGACGGACTCCCCGCCACCTCCATCGCCTGG 12427

Qy      24130 GGCCCCTGGGCCGGAAAGGGCATGTCGGCCGGTGATGCGGCTCATGGTTACCTGGAAAAG 24189

Db      12428 GGCGTGTGGGGCGGCGGCGGTATGGGGGCCGACGAGGCGGGCGCGGAGTATCTGGGCCGG 12487

Qy      24190 CGCGGCATTCTGCCGATGGAGCCACGCATGGCGCTCGCGGCATTCCATCGTGCGCGGGCG 24249

Db      12488 CGCGGTATGCGCCCCATGGCACCGGTCTCCGCGCTCCGGGCGATGGCCACCGCCATCGCC 12547

Qy      24250 CAGCGGCCGAATTCCAACCTGATCATCGCGGACATCGACTGGGAGCGCTTCGTCCCCGCC 24309

Db      12548 TCCGG---GGAACCCTGCCCCACCGTCACCCACACCGACTGGGAGCGCTTCGGCGAGGGC 12604

Qy      24310 TTCACCGCTCGACGCCACAGCCCGCTCATCGAGGACATTCCGGAGGTTCGGCAAGCGGCT 24369

Db      12605 TTCACCGCCTTCCGGCCCAGCCCTCTGATCGCGG----------GGCTCGGCACGCCGGG 12654

Qy      24370 CAGGAGCTGGAAGCAGCTGCGTCGACGGCAAAGACGACCACAGCTCAGCCGATTGCGACG 24429

Db      12655 CGGCGGCCG----------------GGCGGCGGAGACCCCCGAGGAGGGGAACGCCACC 12697

Qy      24430 TCTCTCCGTGAGCGATTGGCCCGACTGACGTCCTCAAAGCAGAACCAGGTGCTGCTCGGC 24489

Db      12698 GCTGCGGCGGACC---TCACCGCCCTGCCGCCCGCCGAACTCCGCACCGCGCTGCGCGAG 12754

Qy      24490 CTGATTCGGACAGGCATCTGCACCGTTCTCGGCCTTCGTAATCCGGAAGGCATCGAGGAC 24549

Db      12755 CTGGTGCGAGCCCGGACCGCCGCGGCGCTCGGCCTCGACGACCCGGCCGAGGTCGCCGAG 12814

Qy      24550 CAACGAGCCTTCCGCGACCTCGGCTTCGACTCGCTGACGTCGGCTCAGTTCAGCAAGGAA 24609

Db      12815 GGCGAACGGTTCCCCGCCATGGGCTTCGACTCCCTGGCCACCGTACGGCTGCGCCGCGGA 12874

Qy      24610 CTCGCCAAGGAAACCGGACTGCCACTCCCCCCGTCCCTGGTCTTCGACTATCCCACCCCG 24669

Db      12875 CTCGCCTCGGCCACGGGCCTCGACCTGCCCCCCGATCTGCTCTTCGACCGGGACACCCCG 12934

Qy      24670 CAGGAATGTGCTGCCCATCTGCGCACACAACTCGTCGACCTAGACGACGAAGAGGACGCG 24729

Db      12935 GCCGCGCTCGCCGCCCACCTGGCCGAACTGCTCGCCACCGCACGGGACCACGGACCCGGC 12994

Qy      24730 GCACTGTCGAATGCTCTCCCGC 24751

Db      12995 GGCCCCGGGACCGGTGCCGCGC 13016
```

RESULT 15
US-09-860-846-32
; Sequence 32, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-32

Query Match 7.0%; Score 2140.6; DB 9; Length 11220;
Best Local Similarity 54.0%; Pred. No. 0;
Matches 5775; Conservative 0; Mismatches 4324; Indels 600; Gaps 44;

```
Qy      19950 CGCGGAGTTCCAGCCCGACAACTCAACTCCGCTTCCGCGACCCCGGGCAGAGCTCGATGA 20009

Db         48 CACGGCGGACCTCCACGAGGCCCGTGGCCGCCTCCGCGAGCTGGAGGCGAAGGCGGGCGA 107

Qy      20010 GCCGATCGCCATCGTTGGCATGGCCTGTCGCTTCCCCGGCGGAGTGACCTCGGCGGACGA 20069

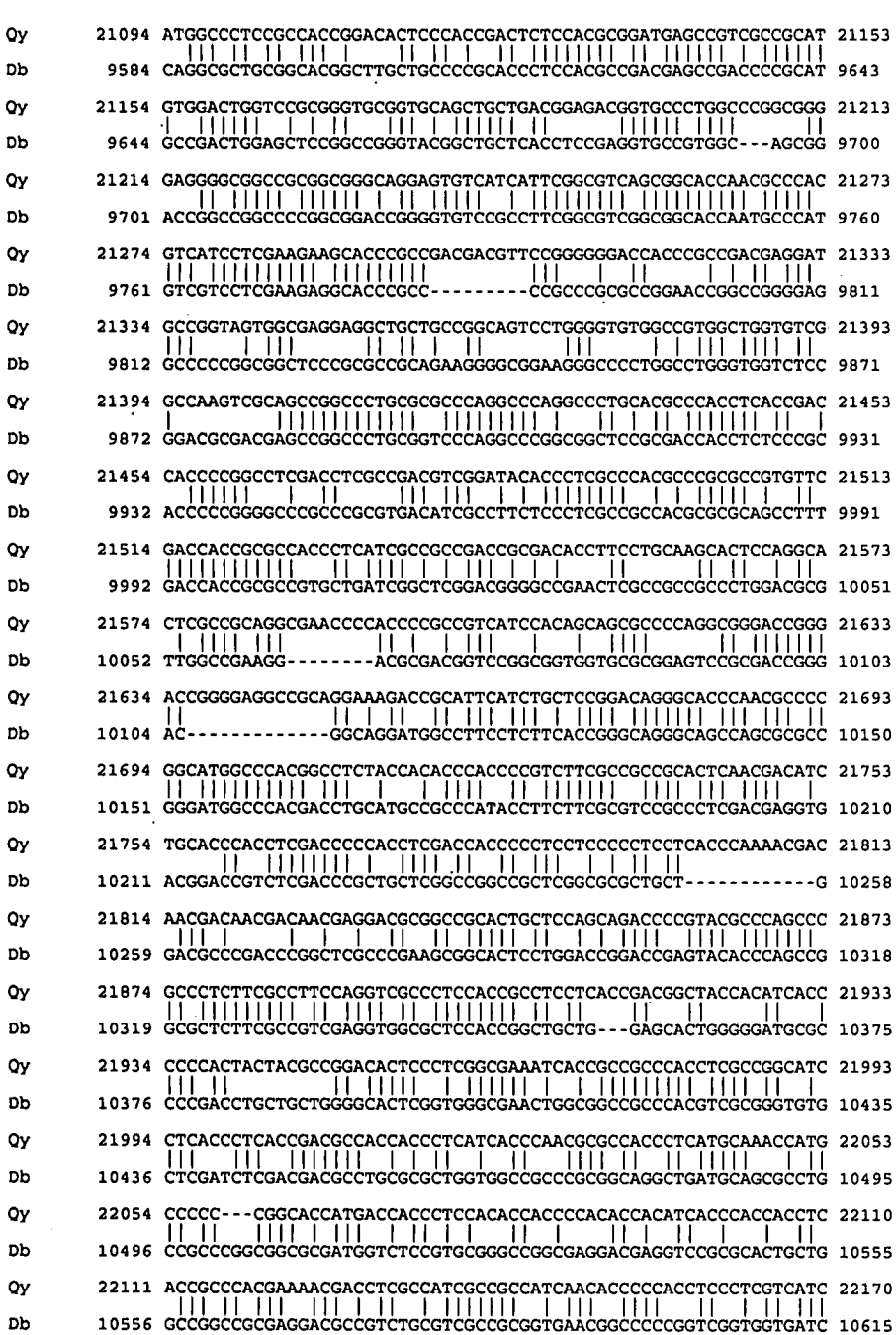
Db        108 GCCGGTGGCGATCGTCGGCATGGCCTGCCGCCTGCCCGGCGGCGTCGCCTCGCCCGAGGA 167

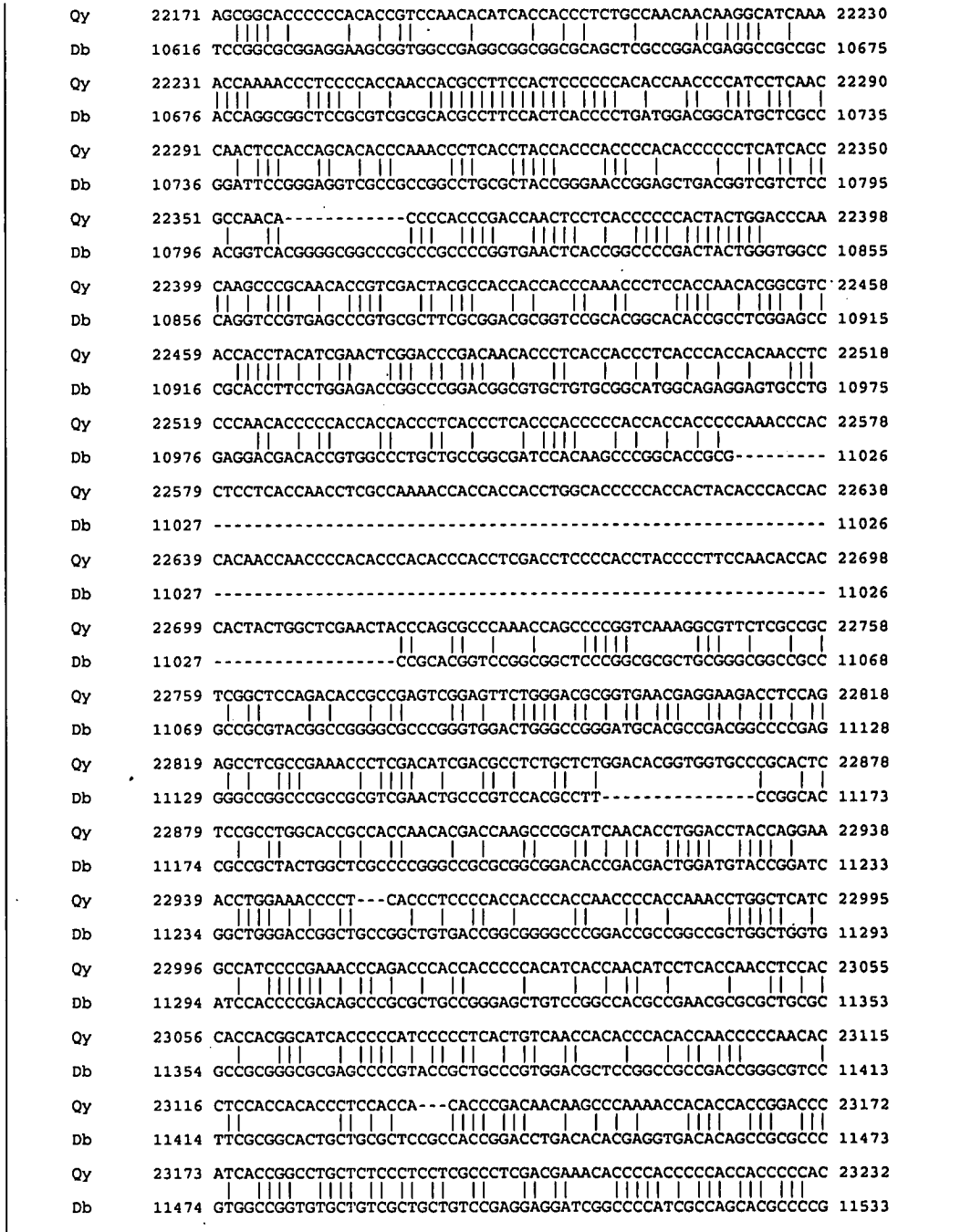
Qy      20070 CTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGCGG 20129

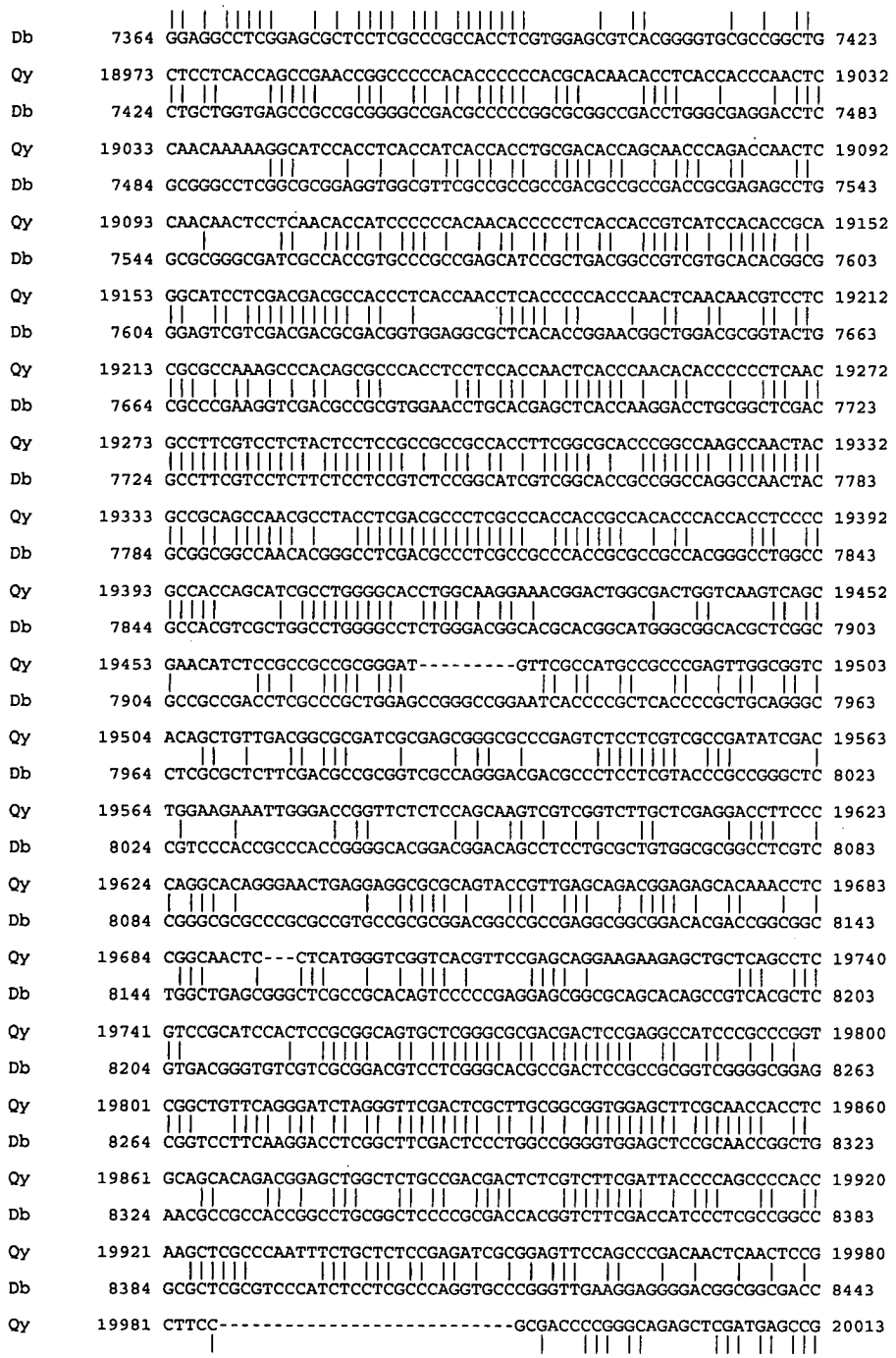
Db        168 CCTGTGGCGGCTGGTGGCCGGCGGCGAGGACGCGATCTCGGAGTTCCCCCAGGACCGCGG 227

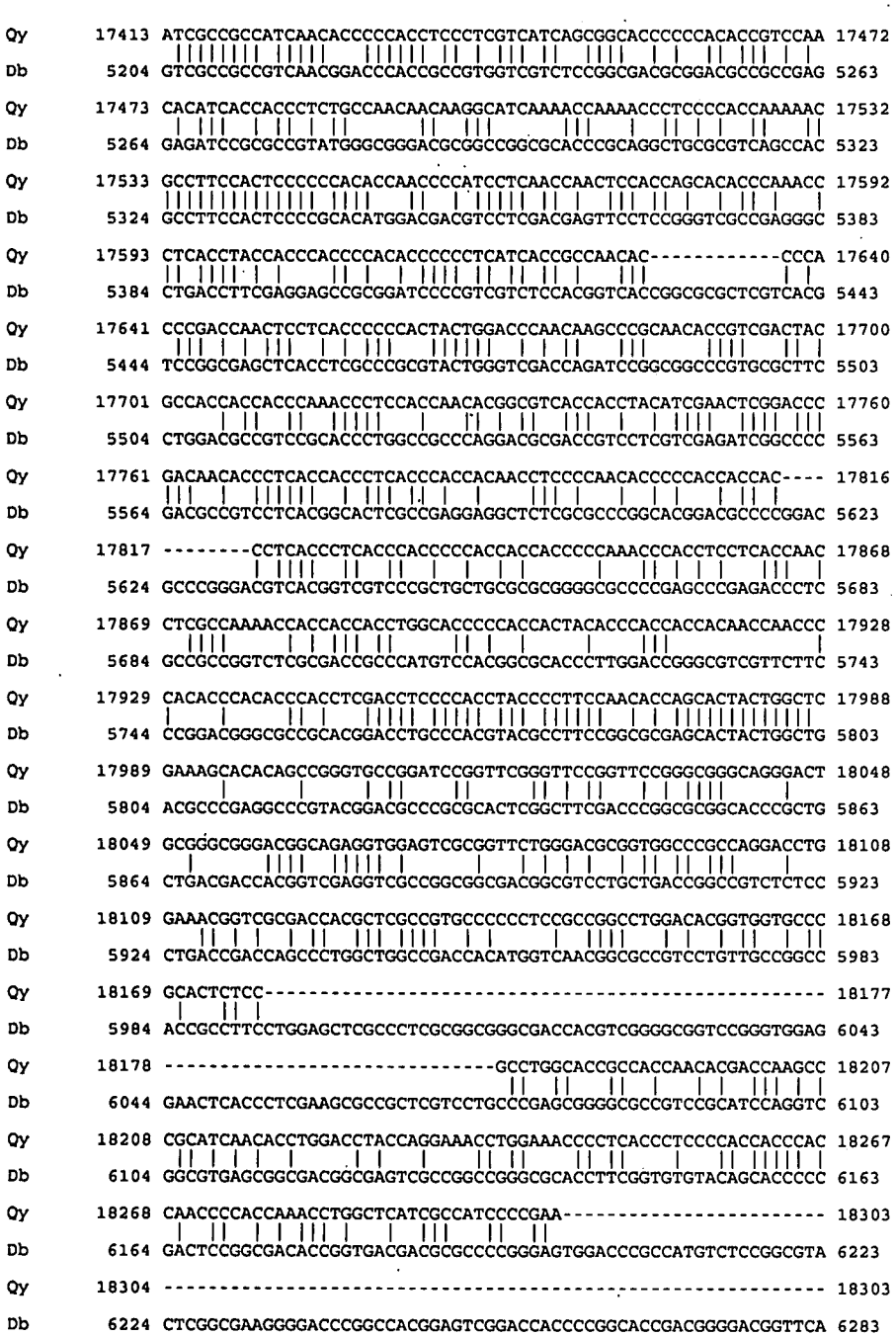
Qy      20130 CTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACCCG 20189

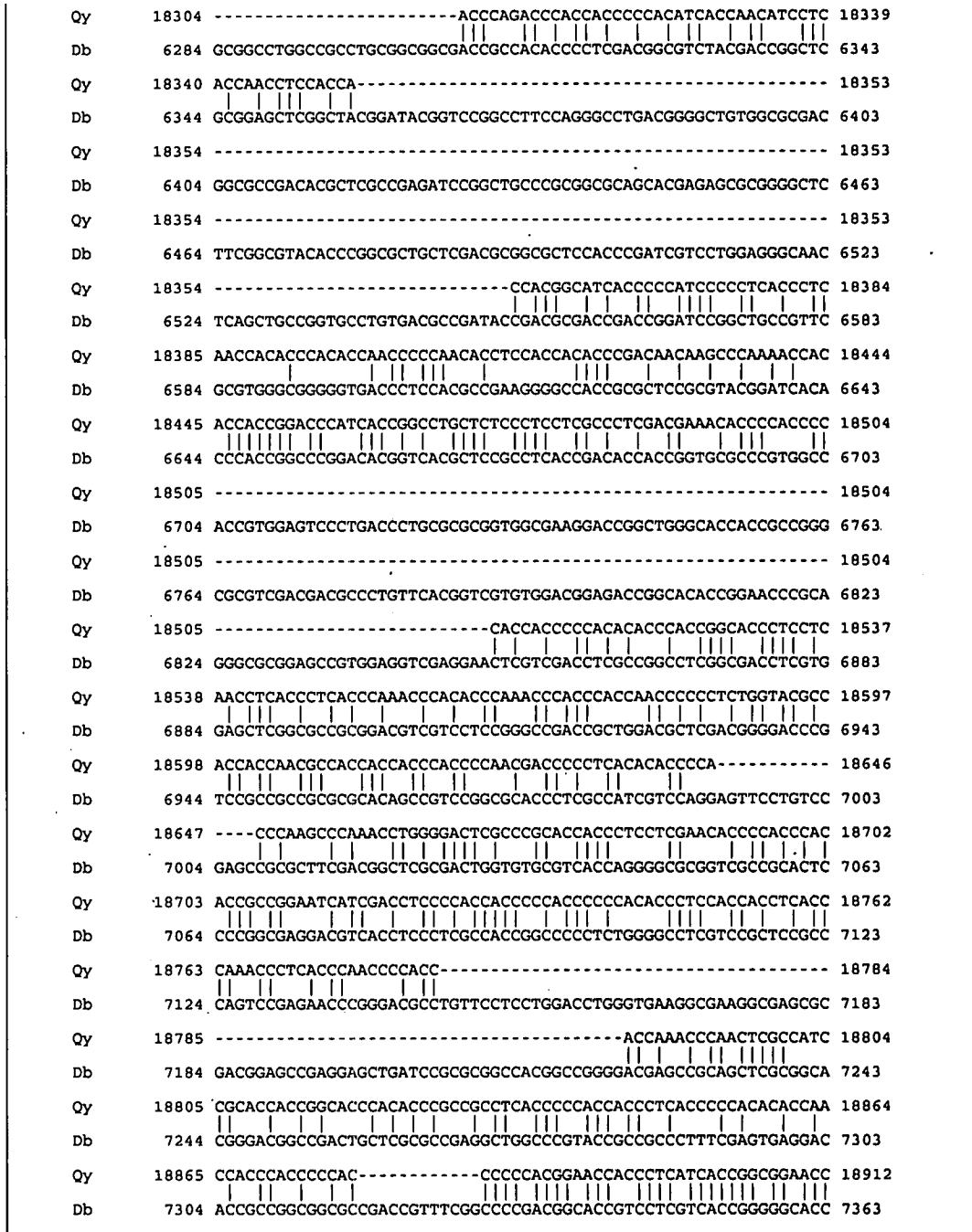
Db        228 CTGGGACGTGGAGGGCCTGTACGACCCGAACCCGGAGGCCACGGGCAAGAGTTACGCCCG 287
```

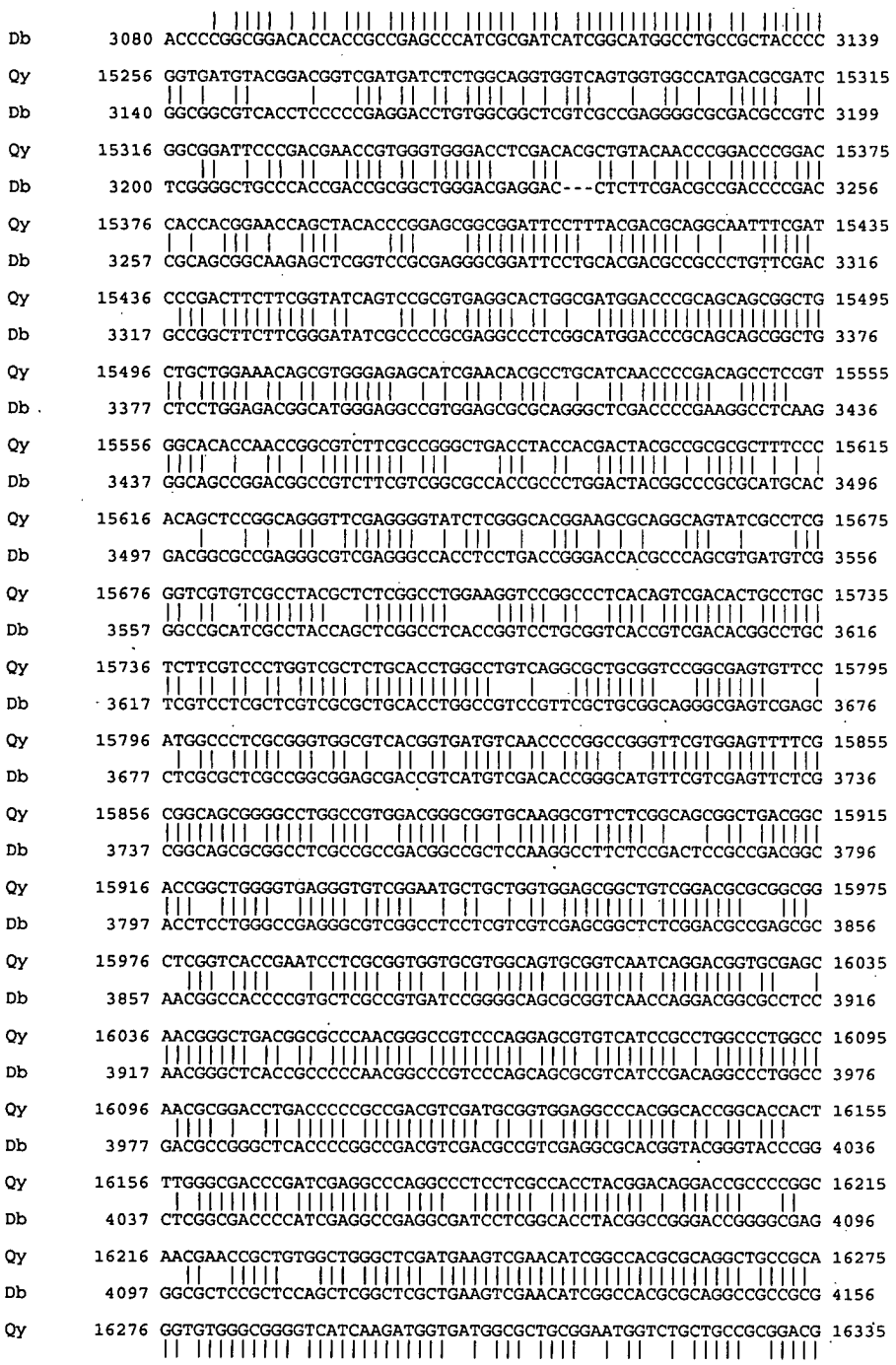

Qy 21094 ATGGCCCTCCGCCACCGGACACTCCCACCGACTCTCCACGCGGATGAGCCGTCGCCGCAT 21153
Db 9584 CAGGCGCTGCGGCACGGCTTGCTGCCCCGCACCCTCCACGCCGACGAGCCGACCCCGCAT 9643
Qy 21154 GTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCCGGCGGG 21213
Db 9644 GCCGACTGGAGCTCCGGCCGGGTACGGCTGCTCACCTCCGAGGTGCCGTGGC---AGCGG 9700
Qy 21214 GAGGGGCGGCCGCGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAACGCCCAC 21273
Db 9701 ACCGGCCGGCCCCGGCGGACCGGGGTGTCCGCCTTCGGCGTCGGCGGCACCAATGCCCAT 9760
Qy 21274 GTCATCCTCGAAGAAGCACCCGCCGACGACGTTCCGGGGGGACCACCCGCCGACGAGGAT 21333
Db 9761 GTCGTCCTCGAAGAGGCACCCGCC---------CCGCCCGCGCCGGAACCGGCCGGGGAG 9811
Qy 21334 GCCGGTAGTGGCGAGGAGGCTGCTGCCGGCAGTCCTGGGGTGTGGCCGTGGCTGGTGTCG 21393
Db 9812 GCCCCCGGCGGCTCCCGCGCCGCAGAAGGGGCGGAAGGGCCCCTGGCCTGGGTGGTCTCC 9871
Qy 21394 GCCAAGTCGCAGCCGGCCCTGCGCGCCCAGGCCCAGGCCCTGCACGCCCACCTCACCGAC 21453
Db 9872 GGACGCGACGAGCCGGCCCTGCGGTCCCAGGCCCGGCGGCTCCGCGACCACCTCTCCCGC 9931
Qy 21454 CACCCCGGCCTCGACCTCGCCGACGTCGGATACACCCTCGCCCACGCCCGCGCCGTGTTC 21513
Db 9932 ACCCCCGGGGCCCGCCCGCGTGACATCGCCTTCTCCCTCGCCGCCACGCGCGCAGCCTTT 9991
Qy 21514 GACCACCGCGCCACCCTCATCGCCGCCGACCGCGACACCTTCCTGCAAGCACTCCAGGCA 21573
Db 9992 GACCACCGCGCCGTGCTGATCGGCTCGGACGGGGCCGAACTCGCCGCCGCCCTGGACGCG 10051
Qy 21574 CTCGCCGCAGGCGAACCCCACCCCGCCGTCATCCACAGCAGCGCCCCAGGCGGGACCGGG 21633
Db 10052 TTGGCCGAAGG--------ACGCGACGGTCCGGCGGTGGTGCGCGGAGTCCGCGACCGGG 10103
Qy 21634 ACCGGGGAGGCCGCAGGAAAGACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCC 21693
Db 10104 AC-------------GGCAGGATGGCCTTCCTCTTCACCGGGCAGGGCAGCCAGCGCGCC 10150
Qy 21694 GGCATGGCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATC 21753
Db 10151 GGGATGGCCCACGACCTGCATGCCGCCCATACCTTCTTCGCGTCCGCCCTCGACGAGGTG 10210
Qy 21754 TGCACCCACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGAC 21813
Db 10211 ACGGACCGTCTCGACCCGCTGCTCGGCCGGCCGCTCGGCGCGCTGCT------------G 10258
Qy 21814 AACGACAACGACAACGAGGACGCGGCCGCACTGCTCCAGCAGACCCCGTACGCCCAGCCC 21873
Db 10259 GACGCCCGACCCGGCTCGCCCGAAGCGGCACTCCTGGACCGGACCGAGTACACCCAGCCG 10318
Qy 21874 GCCCTCTTCGCCTTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACC 21933
Db 10319 GCGCTCTTCGCCGTCGAGGTGGCGCTCCACCGGCTGCTG---GAGCACTGGGGGATGCGC 10375
Qy 21934 CCCCACTACTACGCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATC 21993
Db 10376 CCCGACCTGCTGCTGGGGCACTCGGTGGGCGAACTGGCGGCCGCCCACGTCGCGGGTGTG 10435
Qy 21994 CTCACCCTCACCGACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATG 22053
Db 10436 CTCGATCTCGACGACGCCTGCGCGCTGGTGGCCGCCCGCGGCAGGCTGATGCAGCGCCTG 10495
Qy 22054 CCCCC---CGGCACCATGACCACCCTCCACACCACCCCACACCACATCACCCACCACCTC 22110
Db 10496 CCGCCCGGCGGCGCGATGGTCTCCGTGCGGGCCGGCGAGGACGAGGTCCGCGCACTGCTG 10555
Qy 22111 ACCGCCCACGAAAACGACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATC 22170
Db 10556 GCCGGCCGCGAGGACGCCGTCTGCGTCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATC 10615

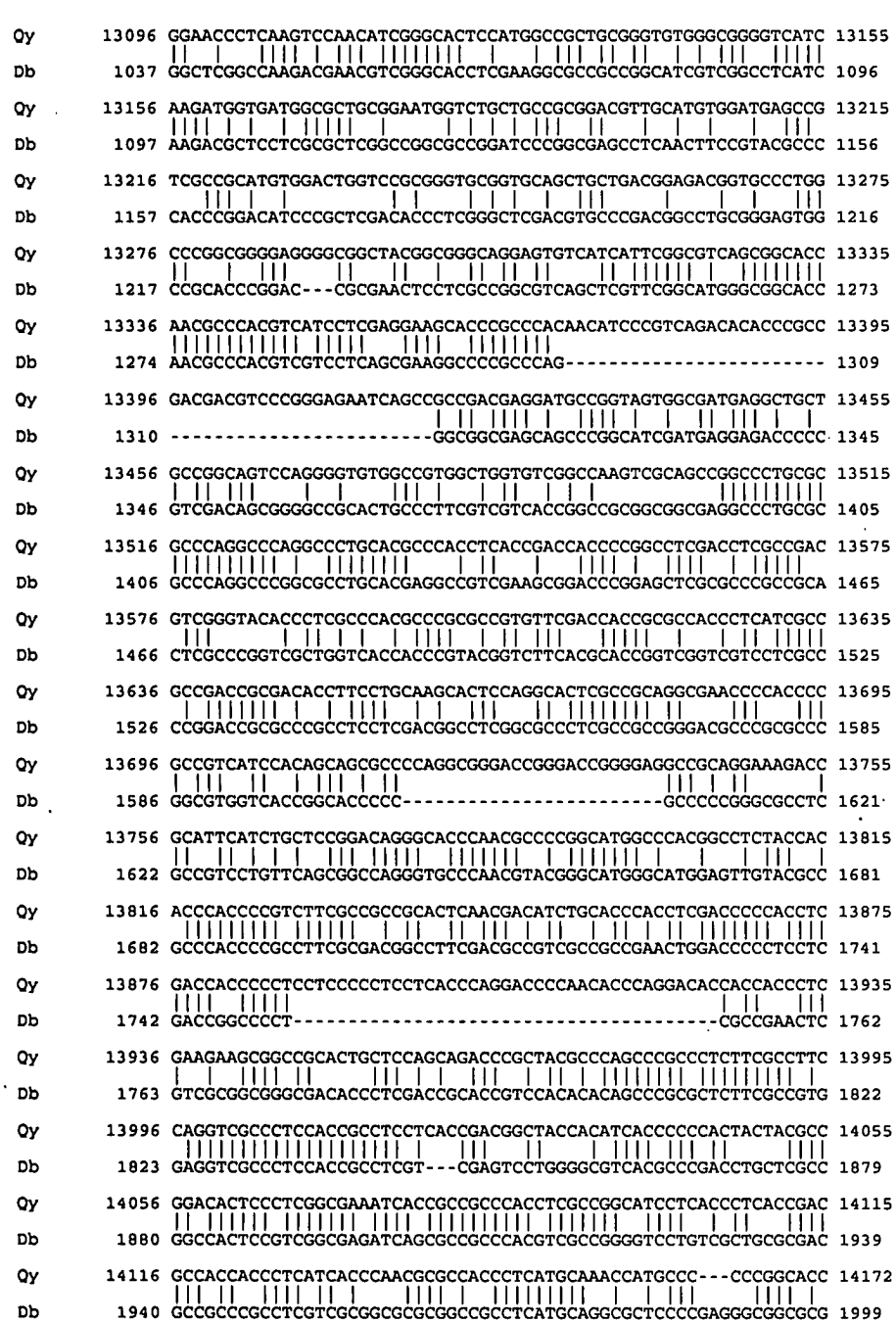

Qy 22171 AGCGGCACCCCCCACACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAA 22230
Db 10616 TCCGGCGCGGAGGAAGCGGTGGCCGAGGCGGCGGCGCAGCTCGCCGGACGAGGCCGCCGC 10675
Qy 22231 ACCAAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAAC 22290
Db 10676 ACCAGGCGGCTCCGCGTCGCGCACGCCTTCCACTCACCCCTGATGGACGGCATGCTCGCC 10735
Qy 22291 CAACTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACC 22350
Db 10736 GGATTCCGGGAGGTCGCCGCGCCGGCCTGCGCTACCGGGAACCGGAGCTGACGGTCGTCTCC 10795
Qy 22351 GCCAACA----------CCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAA 22398
Db 10796 ACGGTCACGGGGCGGCCCGCCCGCCCCGGTGAACTCACCGGCCCCGACTACTGGGTGGCC 10855
Qy 22399 CAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTC 22458
Db 10856 CAGGTCCGTGAGCCCGTGCGCTTCGCGGACGCGGTCCGCACGGCACACCGCCTCGGAGCC 10915
Qy 22459 ACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTC 22518
Db 10916 CGCACCTTCCTGGAGACCGGCCCGGACGGCGTGCTGTGCGGCATGGCAGAGGAGTGCCTG 10975
Qy 22519 CCCAACACCCCCACCACCACCCTCACCCTCACCCACCCCCACCACCACCCCCAAACCCAC 22578
Db 10976 GAGGACGACACCGTGGCCCTGCTGCCGGCGATCCACAAGCCCGGCACCGCG--------- 11026
Qy 22579 CTCCTCACCAACCTCGCCAAAACCACCACCACCTGGCACCCCCACCACTACACCCACCAC 22638
Db 11027 ---------- 11026
Qy 22639 CACAACCAACCCCACACCCACACCCACCTCGACCTCCCCACCTACCCCTTCCAACACCAC 22698
Db 11027 ---------- 11026
Qy 22699 CACTACTGGCTCGAACTACCCAGCGCCCAAACCAGCCCCGGTCAAAGGCGTTCTCGCCGC 22758
Db 11027 ----------CCGCACGGTCCGGCGGCTCCCGGCGCGCTGCGGGCGGCCGCC 11068
Qy 22759 TCGGCTCCAGACACCGCCGAGTCGGAGTTCTGGGACGCGGTGAACGAGGAAGACCTCCAG 22818
Db 11069 GCCGCGTACGGCCGGGGCGCCCGGGTGGACTGGGCCGGGATGCACGCCGACGGCCCCGAG 11128
Qy 22819 AGCCTCGCCGAAACCCTCGACATCGACGCCTCTGCTCTGGACACGGTGGTGCCCGCACTC 22878
Db 11129 GGGCCGGCCCGCCGCGTCGAACTGCCCGTCCACGCCTT--------------CCGGCAC 11173
Qy 22879 TCCGCCTGGCACCGCCACCAACACGACCAAGCCCGCATCAACACCTGGACCTACCAGGAA 22938
Db 11174 CGCCGCTACTGGCTCGCCCCGGGCCGCGCGGCGGACACCGACGACTGGATGTACCGGATC 11233
Qy 22939 ACCTGGAAACCCCT---CACCCTCCCCACCACCCACCAACCCCACCAAACCTGGCTCATC 22995
Db 11234 GGCTGGGACCGGCTGCCGGCTGTGACCGGCGGGGCCCGGACCGCCGGCCGCTGGCTGGTG 11293
Qy 22996 GCCATCCCCGAAACCCAGACCCACCACCCCCACATCACCAACATCCTCACCAACCTCCAC 23055
Db 11294 ATCCACCCCGACAGCCCGCGCTGCCGGGAGCTGTCCGGCCACGCCGAACGCGCGCTGCGC 11353
Qy 23056 CACCACGGCATCACCCCCATCCCCCTCACTGTCAACCACACCCACACCAACCCCCAACAC 23115
Db 11354 GCCGCGGGCGCGAGCCCCGTACCGCTGCCCGTGGACGCTCCGGCCGCCGACCGGGCGTCC 11413
Qy 23116 CTCCACCACACCCTCCACCA---CACCCGACAACAAGCCCAAAACCACACCACCGGACCC 23172
Db 11414 TTCGCGGCACTGCTGCGCTCCGCCACCGGACCTGACACACGAGGTGACACAGCCGCGCCC 11473
Qy 23173 ATCACCGGCCTGCTCTCCCTCCTCGCCCTCGACGAAACACCCCACCCCCACCACCCCCAC 23232
Db 11474 GTGGCCGGTGTGCTGTCGCTGCTGTCCGAGGAGGATCGGCCCCATCGCCAGCACGCCCCG 11533
Qy 23233 ACACCCACCGGCACCCTCCTCAACCTCACCCTCCCCCAAACCCACACCCAAACCCACCCA 23292

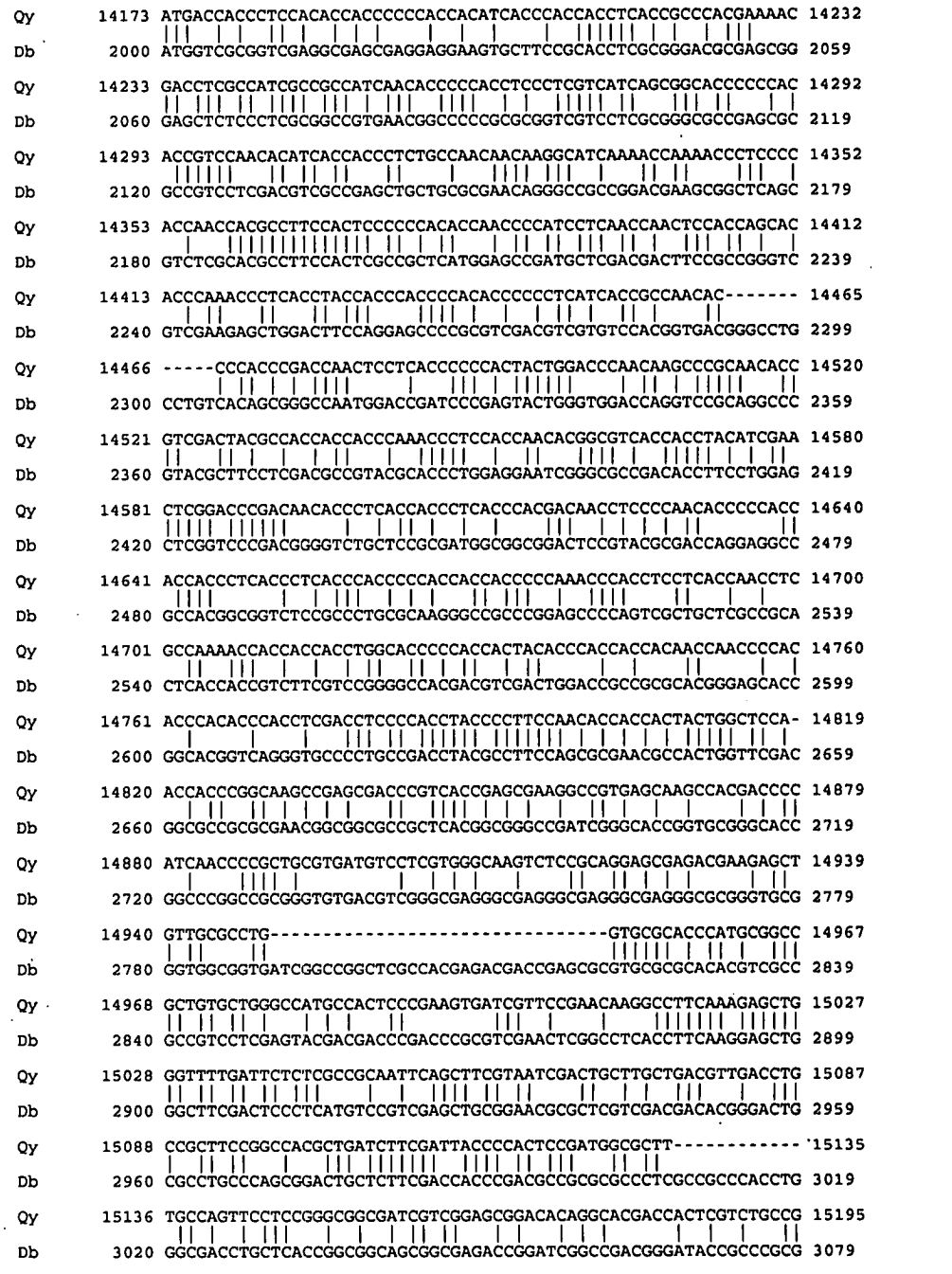

Db 7364 GGAGGCCTCGGAGCGCTCCTCGCCCGCCACCTCGTGGAGCGTCACGGGGTGCGCCGGCTG 7423
Qy 18973 CTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACGCACAACACCTCACCACCCAACTC 19032
Db 7424 CTGCTGGTGAGCCGCCGCGGGGCCGACGCCCCCGGCGCGGCCGACCTGGGCGAGGACCTC 7483
Qy 19033 CAACAAAAAGGCATCCACCTCACCATCACCACCTGCGACACCAGCAACCCAGACCAACTC 19092
Db 7484 GCGGGCCTCGGCGCGGAGGTGGCGTTCGCCGCCGCCGACGCCGCCGACCGCGAGAGCCTG 7543
Qy 19093 CAACAACTCCTCAACACCATCCCCCCACAACACCCCCTCACCACCGTCATCCACACCGCA 19152
Db 7544 GCGCGGGCGATCGCCACCGTGCCCCGCCGAGCATCCGCTGACGGCCGTCGTGCACACGGCG 7603
Qy 19153 GGCATCCTCGACGACGCCACCCTCACCAACCTCACCCCCACCCAACTCAACAACGTCCTC 19212
Db 7604 GGAGTCGTCGACGACGCGACGGTGGAGGCGCTCACACCGGAACGGCTGGACGCGGTACTG 7663
Qy 19213 CGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAACTCACCCAACACACCCCCCTCAAC 19272
Db 7664 CGCCCGAAGGTCGACGCCGCGTGGAACCTGCACGAGCTCACCAAGGACCTGCGGCTCGAC 7723
Qy 19273 GCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCGGCGCACCCGGCCAAGCCAACTAC 19332
Db 7724 GCCTTCGTCCTCTTCTCCTCCGTCTCCGGCATCGTCGGCACCGCCGGCCAGGCCAACTAC 7783
Qy 19333 GCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACCACCGCCACACCCACCACCTCCCC 19392
Db 7784 GCGGCGGCCAACACGGGCCTCGACGCCCTCGCCGCCCACCGCGCCGCCACGGGCCTGGCC 7843
Qy 19393 GCCACCAGCATCGCCTGGGGCACCTGGCAAGGAAACGGACTGGCGACTGGTCAAGTCAGC 19452
Db 7844 GCCACGTCGCTGGCCTGGGGCCTCTGGGACGGCACGCACGGCATGGGCGGCACGCTCGGC 7903
Qy 19453 GAACATCTCCGCCGCCGCGGGAT---------GTTCGCCATGCCGCCCGAGTTGGCGGTC 19503
Db 7904 GCCGCCGACCTCGCCCGCTGGAGCCGGGCCGGAATCACCCCGCTCACCCCGCTGCAGGGC 7963
Qy 19504 ACAGCTGTTGACGGCGCGATCGCGAGCGGGCGCCCGAGTCTCCTCGTCGCCGATATCGAC 19563
Db 7964 CTCGCGCTCTTCGACGCCGCGGTCGCCAGGGACGACGCCCTCCTCGTACCCGCCGGGCTC 8023
Qy 19564 TGGAAGAAATTGGGACCGGTTCTCTCCAGCAAGTCGTCGGTCTTGCTCGAGGACCTTCCC 19623
Db 8024 CGTCCCACCGCCCACCGGGGCACGGACGGACAGCCTCCTGCGCTGTGGCGCGGCCTCGTC 8083
Qy 19624 CAGGCACAGGGAACTGAGGAGGCGCGCAGTACCGTTGAGCAGACGGAGAGCACAAACCTC 19683
Db 8084 CGGGCGCGCCCGCGCCGTGCCGCGCGGACGGCCGCCGAGGCGGCGGACACGACCGGCGGC 8143
Qy 19684 CGGCAACTC---CTCATGGGTCGGTCACGTTCCGAGCAGGAAGAAGAGCTGCTCAGCCTC 19740
Db 8144 TGGCTGAGCGGGCTCGCCGCACAGTCCCCCGAGGAGCGGCGCAGCACAGCCGTCACGCTC 8203
Qy 19741 GTCCGCATCCACTCCGCGGCAGTGCTCGGGCGCGACGACTCCGAGGCCATCCCGCCCGGT 19800
Db 8204 GTGACGGGTGTCGTCGCGGACGTCCTCGGGCACGCCGACTCCGCCGCGGTCGGGGCGGAG 8263
Qy 19801 CGGCTGTTCAGGGATCTAGGGTTCGACTCGCTTGCGGCGGTGGAGCTTCGCAACCACCTC 19860
Db 8264 CGGTCCTTCAAGGACCTCGGCTTCGACTCCCTGGCCGGGGTGGAGCTCCGCAACCGGCTG 8323
Qy 19861 GCAGCACAGACGGAGCTGGCTCTGCCGACGACTCTCGTCTTCGATTACCCCAGCCCCACC 19920
Db 8324 AACGCCGCCACCGGCCTGCGGCTCCCCGCGACCACGGTCTTCGACCATCCCTCGCCGGCC 8383
Qy 19921 AAGCTCGCCCAATTTCTGCTCTCCGAGATCGCGGAGTTCCAGCCCGACAACTCAACTCCG 19980
Db 8384 GCGCTCGCGTCCCATCTCCTCGCCCAGGTGCCCGGGTTGAAGGAGGGGACGGCGGCGACC 8443
Qy 19981 CTTCC----------GCGACCCCGGGCAGAGCTCGATGAGCCG 20013
Db 8444 GCGACCGTCGTGGCCGAGCGGGGCGCTTCCTTCGGTGACCGTGCGACCGACGACGATCCG 8503
Qy 20014 ATCGCCATCGTTGGCATGGCCTGTCGCTTCCCCGGCGGAGTGACCTCGGCGGACGACTTC 20073
Db 8504 ATCGCGATCGTGGGCATGGCATGCCGCTATCCGGGTGGTGTGTCGTCGCCGGAGGACCTG 8563
Qy 20074 TGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGCGGCTGG 20133
Db 8564 TGGCGGCTGGTGGCCGAGGGGACGGACGCGATCAGCGAGTTCCCCGTCAACCGCGGCTGG 8623
Qy 20134 GACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACCCGAAAC 20193
Db 8624 GACCTGGAGAGCCTCTACGACCCGGATCCCGAGTCGAAGGGCACCACGTACTGCCGGGAG 8683
Qy 20194 GGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGCCCCCGC 20253
Db 8684 GGCGGGTTCCTGGAAGGCGCCGGTGACTTCGACGCCGCCTTCTTCGGCATCTCGCCGCGC 8743
Qy 20254 GAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAAACCATC 20313
Db 8744 GAGGCCCTGGTGATGGACCCGCAGCAGCGGCTGCTGCTGGAGGTGTCCTGGGAGGCGCTG 8803
Qy 20314 GAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTCACCGGC 20373
Db 8804 GAACGCGCGGGCATCGACCCGTCCTCGCTGCGCGGCAGCCGCGGTGGTGTCTACGTGGGC 8863
Qy 20374 ACCAACGGACAGGACCACGCGGCACACATCCGTCAGGCCCCGAGCGGTACCGAGGGATTC 20433
Db 8864 GCCGCGCACGGCTCGTACGCCTCCGATCCCCGGCTGGTGCCCGAGGGCTCGGAGGGCTAT 8923
Qy 20434 GTCCTGACCGGGGCAGCCACCAGCATCGCCTCCGGCCGAATCTCCTACATCCTCGGGTTG 20493
Db 8924 CTGCTGACCGGCAGCGCCGACGCGGTGATGTCCGGCCGCATCTCCTACGCGCTCGGTCTC 8983
Qy 20494 GAAGGGCCTGCGGTCACCCTCGACACAGCGTGTTCCTCCTCGCTCGTCGCCCTGCACCTC 20553
Db 8984 GAAGGACCGTCCATGACGGTGGAGACGGCCTGCTCCTCCTCGCTGGTGGCGCTGCATCTG 9043
Qy 20554 GCCTGCCAGTCCCTCAGGTCCGGTGAATGCACCATGGCCTTGGCCGGCGGGGCCACGGTC 20613
Db 9044 GCGGTACGGGCGCTGCGGCACGGCGAGTGCGGGCTCGCGCTGGCGGGCGGGGTGGCGGTG 9103
Qy 20614 ATGACCACCCCGATCACCTTCACCGAATTCGCCCGCCAACGCGGACTCGCCCCCGACGGG 20673
Db 9104 ATGGCCGATCCGGCGGCGTTCGTGGAGTTCTCCCGGCAGAAGGGGCTGGCCGCCGACGGC 9163
Qy 20674 CGTTGCAAGGCGTTCTCGGCGGCGGCTGACGGTACCGGCTGGGGTGAGGGTGTGGGGATG 20733
Db 9164 CGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGGCTGGGCCGAGGGCGTCGGCGTG 9223
Qy 20734 CTGCTGGTGGAGCGGCTCTCCGACGCCCGCCGCAACGGTCACCGTGTCCTGGCCGTGGTG 20793
Db 9224 CTCGTCCTGGAGCGGCTGTCGGACGCGCGCCGCGCGGGGCACACGGTCCTCGGCCTGGTC 9283
Qy 20794 CGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAACGGGCCC 20853
Db 9284 ACCGGCACCGCGGTCAACCAGGACGGTGCCTCCAACGGGCTGACCGCGCCCAACGGCCCA 9343
Qy 20854 TCCCAGCAGCGCGTCATCCGCCAGGCCCTCGCCAACGCGGACCTGACCCCCGCCGACGTC 20913
Db 9344 GCCCAGCAACGCGTCATCGCCGAGGCGCTCGCCGACGCCGGGCTGTCCCCGGAGGACGTG 9403
Qy 20914 GATGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCCCAGGCC 20973
Db 9404 GACGCGGTCGAGGCGCACGGCACCGGCACCCGGCTCGGCGACCCCATCGAGGCCGGGGCG 9463
Qy 20974 ATCCTCGCGACCTACGGACAGGACCGTCCCGGCAACGGGCCGTTGTGGCTGGGCTCCGTC 21033
Db 9464 CTGCTCGCCGCCTCCGGACGGAACCGTTCCGGCGACCACCCGCTGTGGCTCGGCTCGCTG 9523
Qy 21034 AAGTCCAACGTCGGACACACACAGGCCGCGGCGGGCGTGGCCGGAGTGATCAAGATGGTG 21093
Db 9524 AAGTCCAACATCGGGCATGCCCAGGCCGCCGCCGGTGTCGGCGGCGTCATCAAGATGCTC 9583






Db 3080 ACCCCGGCGGACACCACCGCCGAGCCCATCGCGATCATCGGCATGGCCTGCCGCTACCCC 3139
Qy 15256 GGTGATGTACGGACGGTCGATGATCTCTGGCAGGTGGTCAGTGGTGGCCATGACGCGATC 15315
Db 3140 GGCGGCGTCACCTCCCCCGAGGACCTGTGGCGGCTCGTCGCCGAGGGGCGCGACGCCGTC 3199
Qy 15316 GGCGGATTCCCGACGAACCGTGGGTGGGACCTCGACACGCTGTACAACCCGGACCCGGAC 15375
Db 3200 TCGGGGCTGCCCACCGACCGCGGCTGGGACGAGGAC---CTCTTCGACGCCGACCCCGAC 3256
Qy 15376 CACCACGGAACCAGCTACACCCGGAGCGGCGGATTCCTTTACGACGCAGGCAATTTCGAT 15435
Db 3257 CGCAGCGGCAAGAGCTCGGTCCGCGAGGGCGGATTCCTGCACGACGCCGCCCTGTTCGAC 3316
Qy 15436 CCCGACTTCTTCGGTATCAGTCCGCGTGAGGCACTGGCGATGGACCCGCAGCAGCGGCTG 15495
Db 3317 GCCGGCTTCTTCGGGATATCGCCCCGCGAGGCCCTCGGCATGGACCCGCAGCAGCGGCTG 3376
Qy 15496 CTGCTGGAAACAGCGTGGGAGAGCATCGAACACGCCTGCATCAACCCCGACAGCCTCCGT 15555
Db 3377 CTCCTGGAGACGGCATGGGAGGCCGTGGAGCGCGCAGGGCTCGACCCCGAAGGCCTCAAG 3436
Qy 15556 GGCACACCAACCGGCGTCTTCGCCGGGCTGACCTACCACGACTACGCCGCGCGCTTTCCC 15615
Db 3437 GGCAGCCGGACGGCCGTCTTCGTCGGCGCCACCGCCCTGGACTACGGCCCGCGCATGCAC 3496
Qy 15616 ACAGCTCCGGCAGGGTTCGAGGGGTATCTCGGGCACGGAAGCGCAGGCAGTATCGCCTCG 15675
Db 3497 GACGGCGCCGAGGGCGTCGAGGGCCACCTCCTGACCGGGACCACGCCCAGCGTGATGTCG 3556
Qy 15676 GGTCGTGTCGCCTACGCTCTCGGCCTGGAAGGTCCGGCCCTCACAGTCGACACTGCCTGC 15735
Db 3557 GGCCGCATCGCCTACCAGCTCGGCCTCACCGGTCCTGCGGTCACCGTCGACACGGCCTGC 3616
Qy 15736 TCTTCGTCCCTGGTCGCTCTGCACCTGGCCTGTCAGGCGCTGCGGTCCGGCGAGTGTTCC 15795
Db 3617 TCGTCCTCGCTCGTCGCGCTGCACCTGGCCGTCCGTTCGCTGCGGCAGGGCGAGTCGAGC 3676
Qy 15796 ATGGCCCTCGCGGGTGGCGTCACGGTGATGTCAACCCCGGCCGGGTTCGTGGAGTTTTCG 15855
Db 3677 CTCGCGCTCGCCGGCGGAGCGACCGTCATGTCGACACCGGGCATGTTCGTCGAGTTCTCG 3736
Qy 15856 CGGCAGCGGGGCCTGGCCGTGGACGGGCGGTGCAAGGCGTTCTCGGCAGCGGCTGACGGC 15915
Db 3737 CGGCAGCGCGGCCTCGCCGCCGACGGCCGCTCCAAGGCCTTCTCCGACTCCGCCGACGGC 3796
Qy 15916 ACCGGCTGGGGTGAGGGTGTCGGAATGCTGCTGGTGGAGCGGCTGTCGGACGCGCGGCGG 15975
Db 3797 ACCTCCTGGGCCGAGGGCGTCGGCCTCCTCGTCGTCGAGCGGCTCTCGGACGCCGAGCGC 3856
Qy 15976 CTCGGTCACCGAATCCTCGCGGTGGTGCGTGGCAGTGCGGTCAATCAGGACGGTGCGAGC 16035
Db 3857 AACGGCCACCCCGTGCTCGCCGTGATCCGGGGCAGCGCGGTCAACCAGGACGGCGCCTCC 3916
Qy 16036 AACGGGCTGACGGCGCCCAACGGGCCGTCCCAGGAGCGTGTCATCCGCCTGGCCCTGGCC 16095
Db 3917 AACGGGCTCACCGCCCCCAACGGCCCGTCCCAGCAGCGCGTCATCCGACAGGCCCTGGCC 3976
Qy 16096 AACGCGGACCTGACCCCCGCCGACGTCGATGCGGTGGAGGCCCACGGCACCGGCACCACT 16155
Db 3977 GACGCCGGGCTCACCCCGGCCGACGTCGACGCCGTCGAGGCGCACGGTACGGGTACCCGG 4036
Qy 16156 TTGGGCGACCCGATCGAGGCCCAGGCCCTCCTCGCCACCTACGGACAGGACCGCCCCGGC 16215
Db 4037 CTCGGCGACCCCATCGAGGCCGAGGCGATCCTCGGCACCTACGGCCGGGACCGGGGCGAG 4096
Qy 16216 AACGAACCGCTGTGGCTGGGCTCGATGAAGTCGAACATCGGCCACGCGCAGGCTGCCGCA 16275
Db 4097 GGCGCTCCGCTCCAGCTCGGCTCGCTGAAGTCGAACATCGGCCACGCGCAGGCCGCCGCG 4156
Qy 16276 GGTGTGGGCGGGGTCATCAAGATGGTGATGGCGCTGCGGAATGGTCTGCTGCCGCGGACG 16335
Db 4157 GGCGTGGGCGGGCTCATCAAGATGGTCCTCGCGATGCGCCACGGCGTCCTGCCCAGGACG 4216
Qy 16336 TTGCATGTGGATGAGCCGTCGCCGCATGTGGACTGGTCCGCGGGGGCGGTGCAGCTGCTG 16395
Db 4217 CTCCACGTGGACCGGCCCACCACCCGCGTCGACTGGGAGGCCGGCGGCGTCGAGCTCCTC 4276
Qy 16396 ACGGAGACGGTGCCCTGGCCCGGCGGGGAGGGGCGGCTGCGGCGGGCAGGAGTGTCATCG 16455
Db 4277 ACCGAGGAGCGGGAGTGGCCGGA---GACGGGCCGCCCGCGCCGCGGCGATCTCCTCC 4333
Qy 16456 TTCGGCGTCAGCGGCACCAACGCCCACGTCATCCTCGAAGAAGCACCCGCCCACAACATC 16515
Db 4334 TTCGGCATCAGCGGCACCAACGCCCACATCGTGGTCGAACAGGCCCCGGAAGCCGGGGAG 4393
Qy 16516 CCGTCAGACACACCCGCCGACGACGCCCCGGGAGAAGCAGCCGCCGACGATGTTCCGGGG 16575
Db 4394 GCGGCGGTCACCACCACCGCCCCGGAAGCAGGGGAAG----------CCGGG 4435
Qy 16576 GAAGCGGCCGGCGACGACGCCGGTACCGGCGGGGAAGCGACTGGTCCTGCTGCCGGCAGT 16635
Db 4436 GAAGCGGCGGACACCACCGCCACCACGACGCCGGCCGCGGTCGG------CGTCCCCGAA 4489
Qy 16636 CCAGGGGTGTGGCCGTGGCTGGTGTCGGCCAAGTCGCAGCCGGCCCTGCGCGCCCAGGCC 16695
Db 4490 CCCGTACGCGCCCCCGTCGTGGTCTCCGCGCGGGACGCCGCCGCCCTGCGCGCCCAGGCC 4549
Qy 16696 CAGGCCCTGCACGCCCACCTCACCGACCACCCCGGCCTCGACCTCGCCGACGTCGGGTAC 16755
Db 4550 GTTCGGCTGCGGACCTTCCTCGACGGCCGACCGGACGTCACCGTCGCCGACCTCGGACGC 4609
Qy 16756 ACCCTCGCCCACGCCCGCGCCGTGTTCGACCACCGCGCCCACCCTCATCGCCGCCGACCGC 16815
Db 4610 TCGCTGGC---CGCCCGTACCGCCTTCGAGCACAAGGCCGCCCTCACCACCGCCACCAGG 4666
Qy 16816 GACACCTTCCTGCAAGCACTCCAGGCACTCGCCGCAGGCGAACCCCACCCCGCCGTCATC 16875
Db 4667 GACGAGCTGCTCGCCGGGCTCGACGCCCTCGGCCGCGGGGAGCAAGCCACGGGCCTGGTC 4726
Qy 16876 CACAGCAGCGCCCCAGGCGGGACCGGGACCGGGGAGGCCGCAGGAAAGACCGCATTCATC 16935
Db 4727 ----------ACCGGCGAACCGGCCAGGGCCGGACGCACGGCCTTCCTG 4765
Qy 16936 TGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCACACCCACCCC 16995
Db 4766 TTCACCGGCCAGGGAGCGCAGCGCGTCGCCATGGGCGAGGAACTGCGCGCCGCGCACCCC 4825
Qy 16996 GTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTCGACCACCCC 17055
Db 4826 GTGTTCGCCGCCGCCCTCGACACCGTGTACGCGGCCCTCGACCGTCACCTCGACCGGCCG 4885
Qy 17056 CTCCTCCCCCTCCTCACCCAGGACCCCAACACCCAGGACACCACCACCCTCGAAGAAGCG 17115
Db 4886 CTGCGGGAGATCGTCGCC----------GCC 4906
Qy 17116 GCCGCACTGCTCCAGCAGACCCCGTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCC 17175
Db 4907 GGGGAGGAGCTGGACCTCACCGCGTACACCCAGCCCGCCCTCTTCGCCTTCGAGGTGGCG 4966
Qy 17176 CTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCC 17235
Db 4967 CTGTTCCGCCTCCT---CGAACACCACGGCCTCGTCCCCGACCTGCTCACCGGCCACTCC 5023
Qy 17236 CTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACC 17295
Db 5024 GTCGGCGAGATCGCCGCCGCGCACGTCGCCGGTGTCCTCTCCCTCGACGACGCCGCACGT 5083
Qy 17296 CTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCC---CCCGGCACCATGACCACC 17352
Db 5084 CTCGTCACCGCCCGCGGCCGGCTCATGCAGTCGGCCCGCGAGGGCGGCGCGATGATCGCC 5143
Qy 17353 CTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCC 17412
Db 5144 GTGCAGGCGGGCGAGGCCGAGGTCGTCGAGTCCCTGAAGGGCTACGAGGGCAGGGTCGCC 5203





```
Db      12128 GAGGACACCCCGCTCACCGCGGTGTTCCACGCCGCGCGGGCGTACCGCAGGTCACGCCGCTG 12187

Qy      23893 TCGGAAACCGATGCCGAATCCTTCTCTTCCGTTACGGCAGCGAAGGCAACGGGCGCGGCG 23952

Db      12188 TCCCGTACCTCGCCCGAGCACTTCGCCGACGTGTACGCGGGCAAGGCGGCGGGCGCCGCG 12247

Qy      23953 ATTCTGCATGAGTTGCTGCTGGACCATGAAAC---GCTTGAACACTTCATTCTCTTCTCG 24009

Db      12248 CACCTGGACGAACTGACCCGCGAACTCGGCGCCGGACTCGACGCGTTCGTCCTCTACTCC 12307

Qy      24010 TCGGGCGCCGGCGCTTGGGGCAGCGGGAATCAGTGCGCATACTCGGCGGCCAACGCATAC 24069

Db      12308 TCCGGCGCCGGCGTCTGGGGCAGCGCCGGCCAGGGTGCCTACGCCGCCGCCAACGCCGCC 12367

Qy      24070 CTGGACGCGCTCGCGACGCATCGTCAGACACATGGACTTCCCGGGGCATCGATCGCCTGG 24129

Db      12368 CTGGACGCGCTCGCCCGGCGCCGTGCGGCGGACGGACTCCCCGCCACCTCCATCGCCTGG 12427

Qy      24130 GGCCCCTGGGCCGGAAAGGGCATGTCGGCCGGTGATGCGGCTCATGGTTACCTGGAAAAG 24189

Db      12428 GGCGTGTGGGGCGGCGGCGGTATGGGGGCCGACGAGGCGGGCGCGGAGTATCTGGGCCGG 12487

Qy      24190 CGCGGCATTCTGCCGATGGAGCCACGCATGGCGCTCGCGGCATTCCATCGTGCGCGGGCG 24249

Db      12488 CGCGGTATGCGCCCCATGGCACCGGTCTCCGCGCTCCGGGCGATGGCCACCGCCATCGCC 12547

Qy      24250 CAGCGGCCGAATTCCAACCTGATCATCGCGGACATCGACTGGGAGCGCTTCGTCCCCGCC 24309

Db      12548 TCCGG---GGAACCCTGCCCCACCGTCACCCACACCGACTGGGAGCGCTTCGGCGAGGGC 12604

Qy      24310 TTCACCGCTCGACGCCACAGCCCGCTCATCGAGGACATTCCGGAGGTTCGGCAAGCGGCT 24369

Db      12605 TTCACCGCCTTCCGGCCCAGCCCTCTGATCGCGG----------GGCTCGGCACGCCGGG 12654

Qy      24370 CAGGAGCTGGAAGCAGCTGCGTCGACGGCAAAGACGACCACAGCTCAGCCGATTGCGACG 24429

Db      12655 CGGCGGCCG----------------GGCGGCGGAGACCCCCGAGGAGGGGAACGCCACC 12697

Qy      24430 TCTCTCCGTGAGCGATTGGCCCGACTGACGTCCTCAAAGCAGAACCAGGTGCTGCTCGGC 24489

Db      12698 GCTGCGGCGGACC---TCACCGCCCTGCCGCCCGCCGAACTCCGCACCGCGCTGCGCGAG 12754

Qy      24490 CTGATTCGGACAGGCATCTGCACCGTTCTCGGCCTTCGTAATCCGGAAGGCATCGAGGAC 24549

Db      12755 CTGGTGCGAGCCCGGACCGCCGCGGCGCTCGGCCTCGACGACCCGGCCGAGGTCGCCGAG 12814

Qy      24550 CAACGAGCCTTCCGCGACCTCGGCTTCGACTCGCTGACGTCGGCTCAGTTCAGCAAGGAA 24609

Db      12815 GGCGAACGGTTCCCCGCCATGGGCTTCGACTCCCTGGCCACCGTACGGCTGCGCCGCGGA 12874

Qy      24610 CTCGCCAAGGAAACCGGACTGCCACTCCCCCCGTCCCTGGTCTTCGACTATCCCACCCCG 24669

Db      12875 CTCGCCTCGGCCACGGGCCTCGACCTGCCCCCCGATCTGCTCTTCGACCGGGACACCCCG 12934

Qy      24670 CAGGAATGTGCTGCCCATCTGCGCACACAACTCGTCGACCTAGACGACGAAGAGGACGCG 24729

Db      12935 GCCGCGCTCGCCGCCCACCTGGCCGAACTGCTCGCCACCGCACGGGACCACGGACCCGGC 12994

Qy      24730 GCACTGTCGAATGCTCTCCCGC 24751

Db      12995 GGCCCCGGGACCGGTGCCGCGC 13016


RESULT 14
US-09-861-289-1
; Sequence 1, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/861,289
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-861-289-1

  Query Match             7.1%;  Score 2188.4;  DB 10;  Length 15872;
  Best Local Similarity   52.5%;  Pred. No. 0;
  Matches 6973;  Conservative    0;  Mismatches 5061;  Indels 1248;  Gaps   50;

Qy      12256 CGCAGCGGAGGGTTTCTCTATGACGCGGGCGACTTCGACGCCGCGTTCTTCGAGTTGTCA 12315

Db        197 CGCCGGGGCGGCTTCCTCGACTCCCTCGACCTCTTCGACGCGGCCTTCTTCGGAATCTCG 256

Qy      12316 CCGCGTGAGGCGCTGGCAATGGACCCGCAGCAGCGCCTGCTGCTCGAAACCACTTGGGAA 12375

Db        257 CCCCGTGAGGCCGCCGCCATGGACCCGCAGCAGCGACTGGTCCTCGAACTCGCCTGGGAG 316

Qy      12376 ACGTTCGAACAGGGCGGAATCGACCCGAGGTCCATGCGCGGAAGCCGGACCGGGGTTTTC 12435

Db        317 GCGCTGGAGGACGCCGGAATCGTCCCCGGCACCCTCGCCGGAAGCCGCACCGCCGTCTTC 376

Qy      12436 GTGGGGATCAATCCGGAGGACTACACCACCGGATACACACATCAGCCCTCAAACGCAGTC 12495

Db        377 GTCGGCACCCTGCGGGACGACTACACGAGCCTCCTCTACCAGCACGGCGAGCAGGCCATC 436

Qy      12496 GAGGGCTACCTGCTCACTGGCAGCGCGGCAAGCATTGCGTCAGGCCGTATCTCCTACAAC 12555

Db        437 ACCCAGCACACCATGGCGGGCGTGAACCGGGGCGTCATCGCCAACCGCGTCTCGTACCAC 496

Qy      12556 TTCGGGCTCGAAGGCCCTGCGATCACTATCGACACCGCGTGTTCCTCCTCGCTCGTCGCC 12615

Db        497 CTCGGCCTGCAGGGCCCGAGCCTCACCGTCGACGCCGCGCAGTCGTCCTCGCTCGTCGCC 556

Qy      12616 CTGCATCTGGCCTGCCAAGCGCTCCGGTCCGGTGAATGCACCATGGCGCTCGCAGGCGGC 12675

Db        557 GTGCACCTGGCCTGCGAGTCCCTGCGCGCCGGGGAGTCCACGACGGCGCTCGTCGCCGGC 616

Qy      12676 GCCTCCGTCATGGCCACTCCCTTCGTCTTCACCGAGTTCTCTCGCCAGCGGGGCCTGGCC 12735

Db        617 GTGAACCTCAACATCCTCGCGGAGAGCGCCGTGACGGAGGAGCGCTTCGGTGGACTCTCC 676

Qy      12736 GCAGACGGCCGGTGCAAGGCGTTTTCGGCGGCGGCGGACGGGACCGGCTGGTCCGAGGGT 12795

Db        677 CCGGACGGCACCGCCTACACCTTCGACGCGCGGGCCAACGGATTCGTCCGGGGCGAGGGC 736

Qy      12796 GTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCGCCGCAACGGTCACCGTGTCCTG 12855

Db        737 GGCGGAGTCGTCGTACTCAAGCCGCTCTCCCGCGCCCTCGCCGACGGCGACCGTGTCCAC 796

Qy      12856 GCCGTCGTCCGCGGCAGCGCCGTCAACCAGGACGGCGCAAGCAACGGCCTGACCGCACCC 12915

Db        797 GGCGTCATCCGCGCCAGCGCCGTCAACAACGACGGAGCCACCCCGGGTCTCACCGTGCCC 856

Qy      12916 AACGGTCGTTCACAAGTCAAGGTCATCCGCCAGGCTTTGGCCAACGCACACCTCTCCCCT 12975

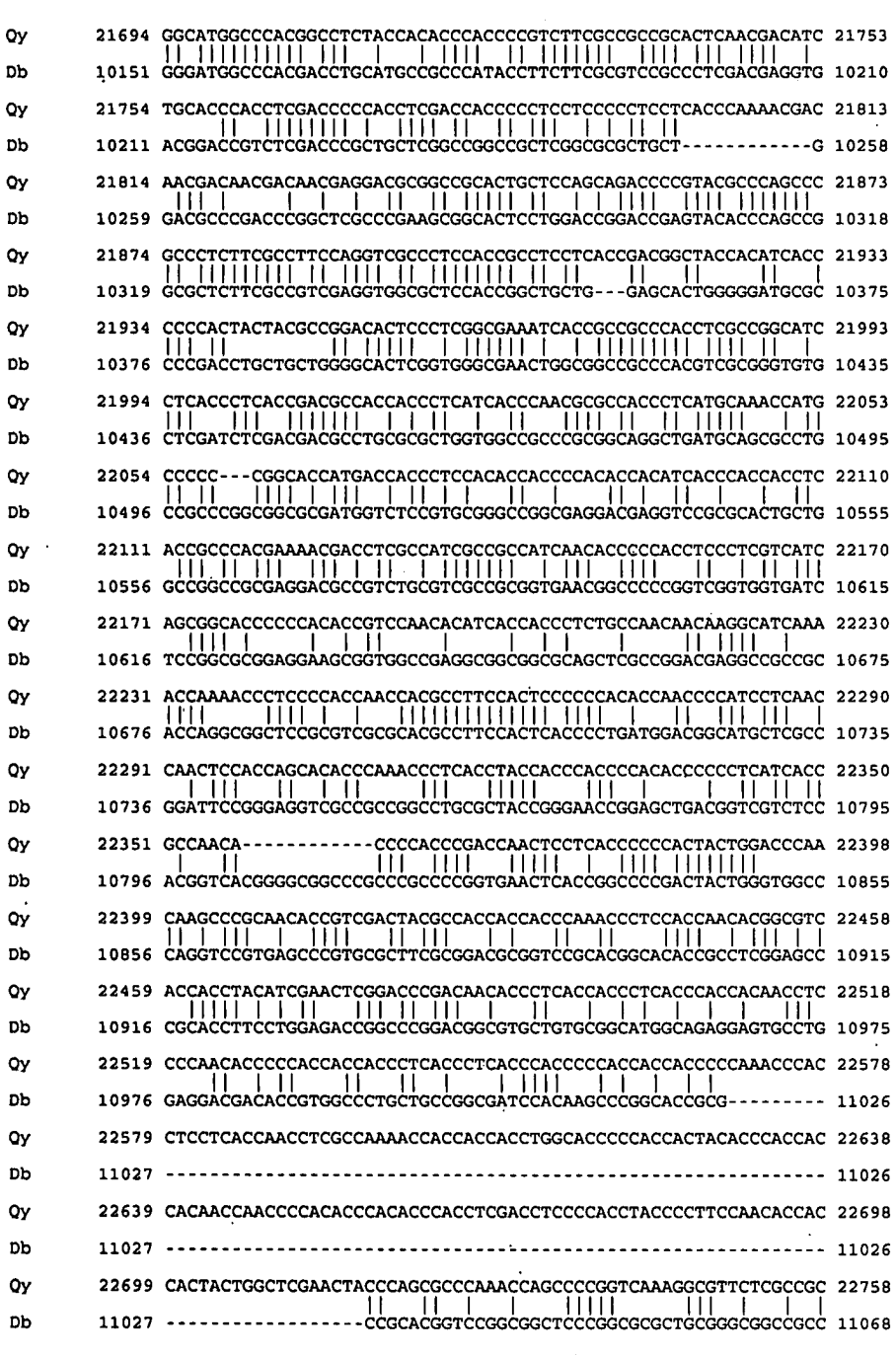
Db        857 AGCAGGGCCGCCCAGGAGAAGGTGCTGCGCGAGGCGTACCGGAAGGCGGCCCTGGACCCG 916

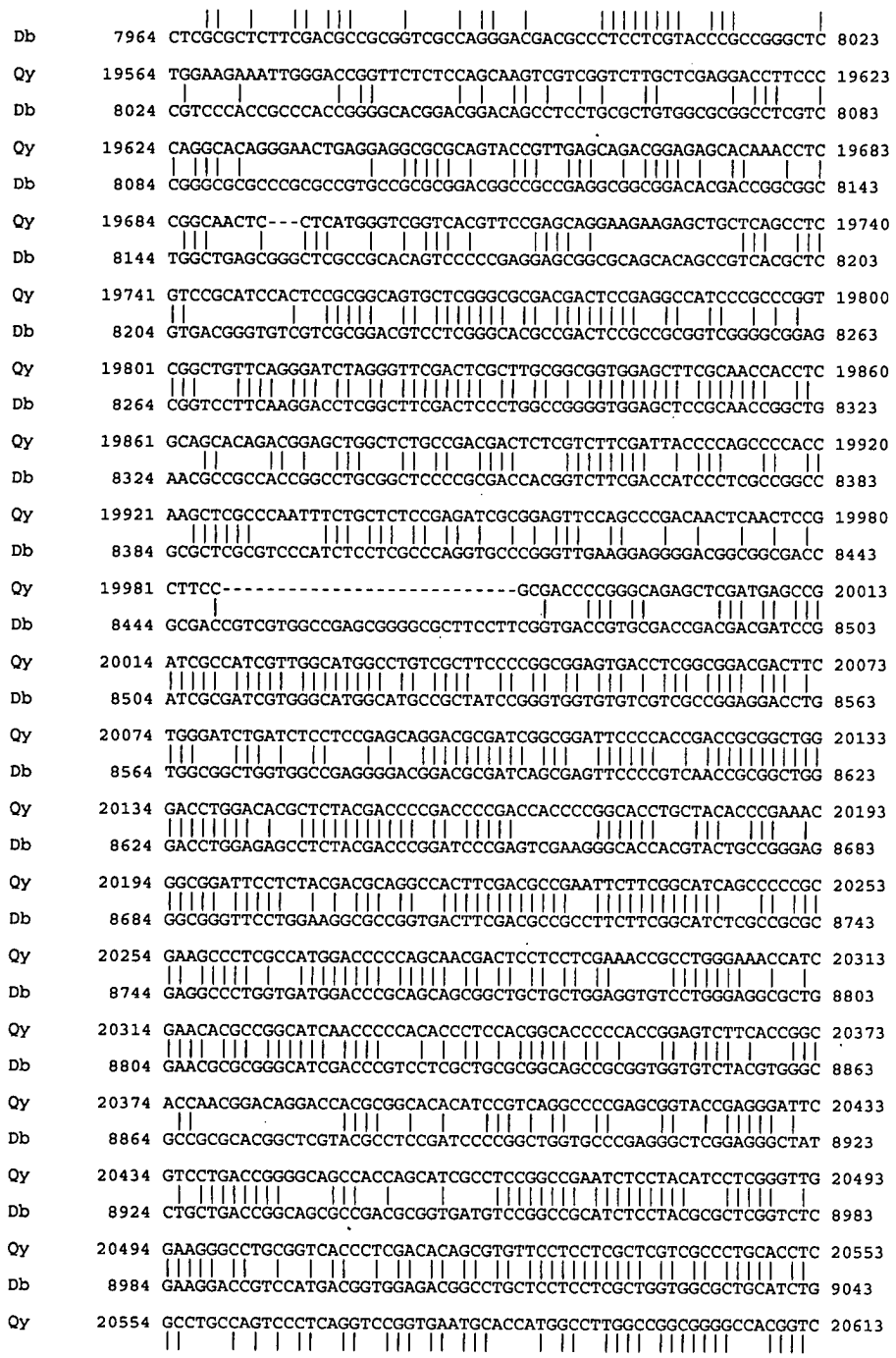
Qy      12976 GCCGATGTCGATGCGGTGGAGGCCCACGGCACGGGGACCACCCTGGGCGACCCGATCGAG 13035

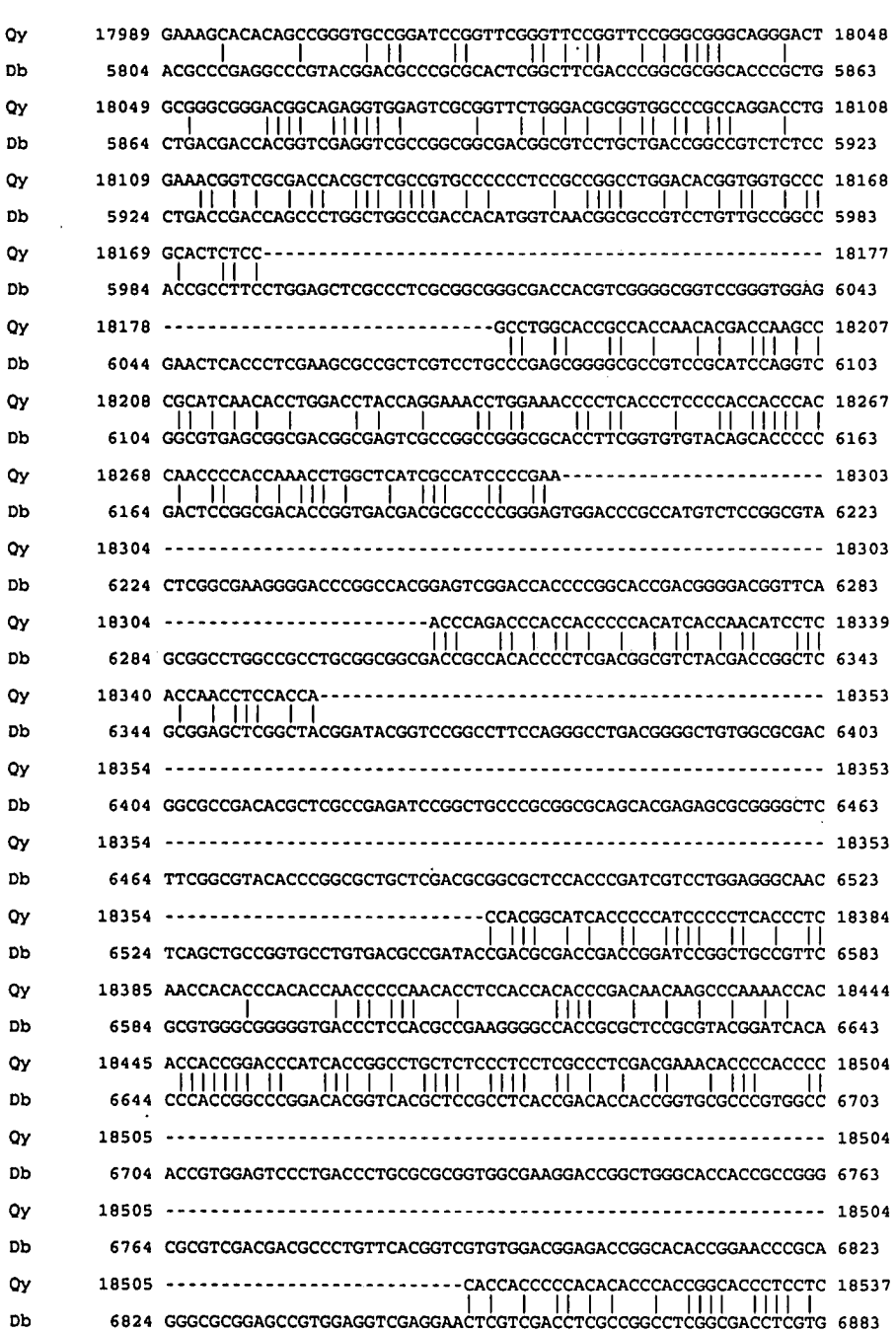
Db        917 TCCGCCGTCCAGTACGTCGAACTCCACGGCACCGGAACCCCCGTCGGCGACCCCATCGAG 976

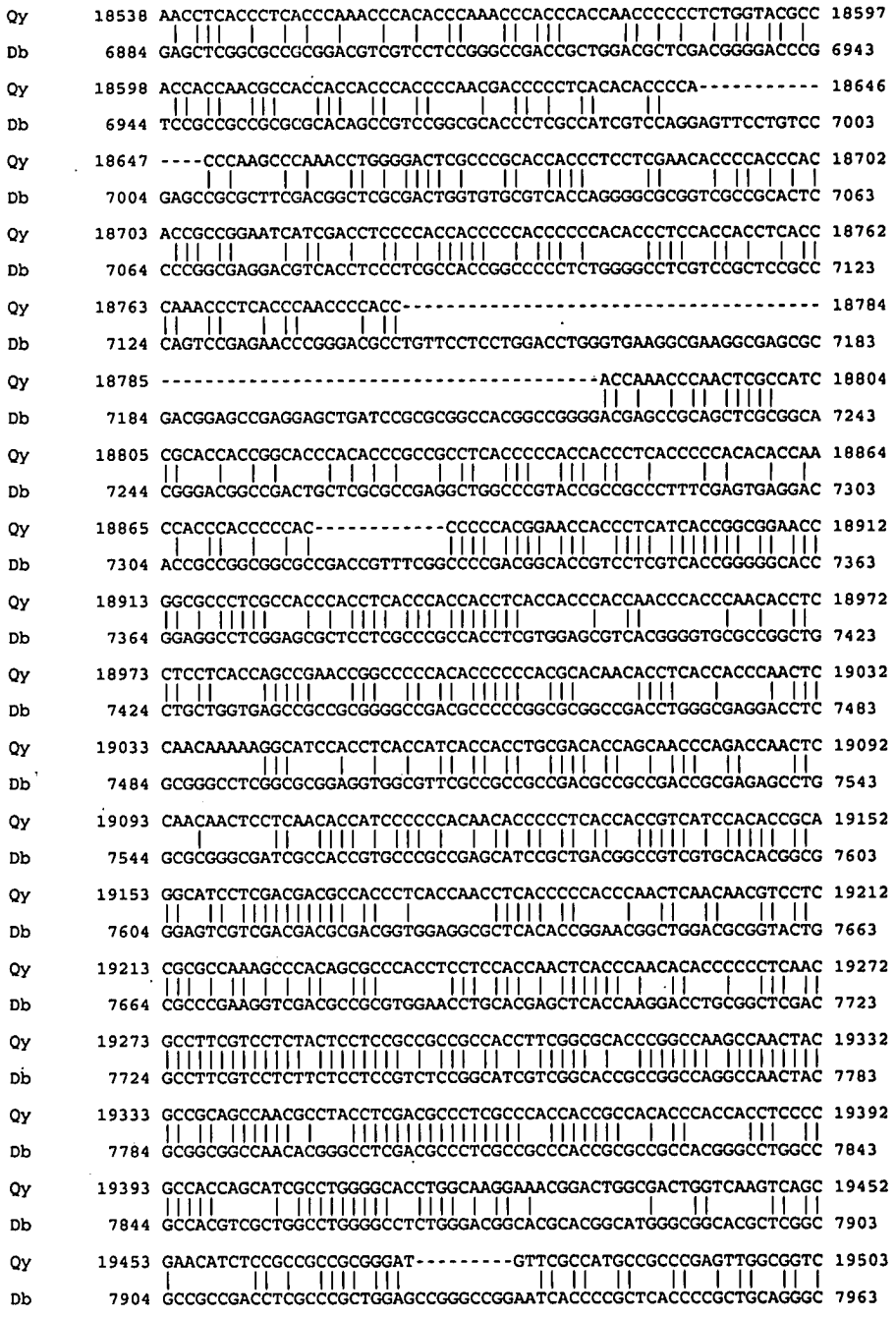
Qy      13036 GCTCAAGCCCTCGTCGAAGCCTACGGTCAGGACCGCCCCAACGGCCGCCCCCTCTGGCTC 13095

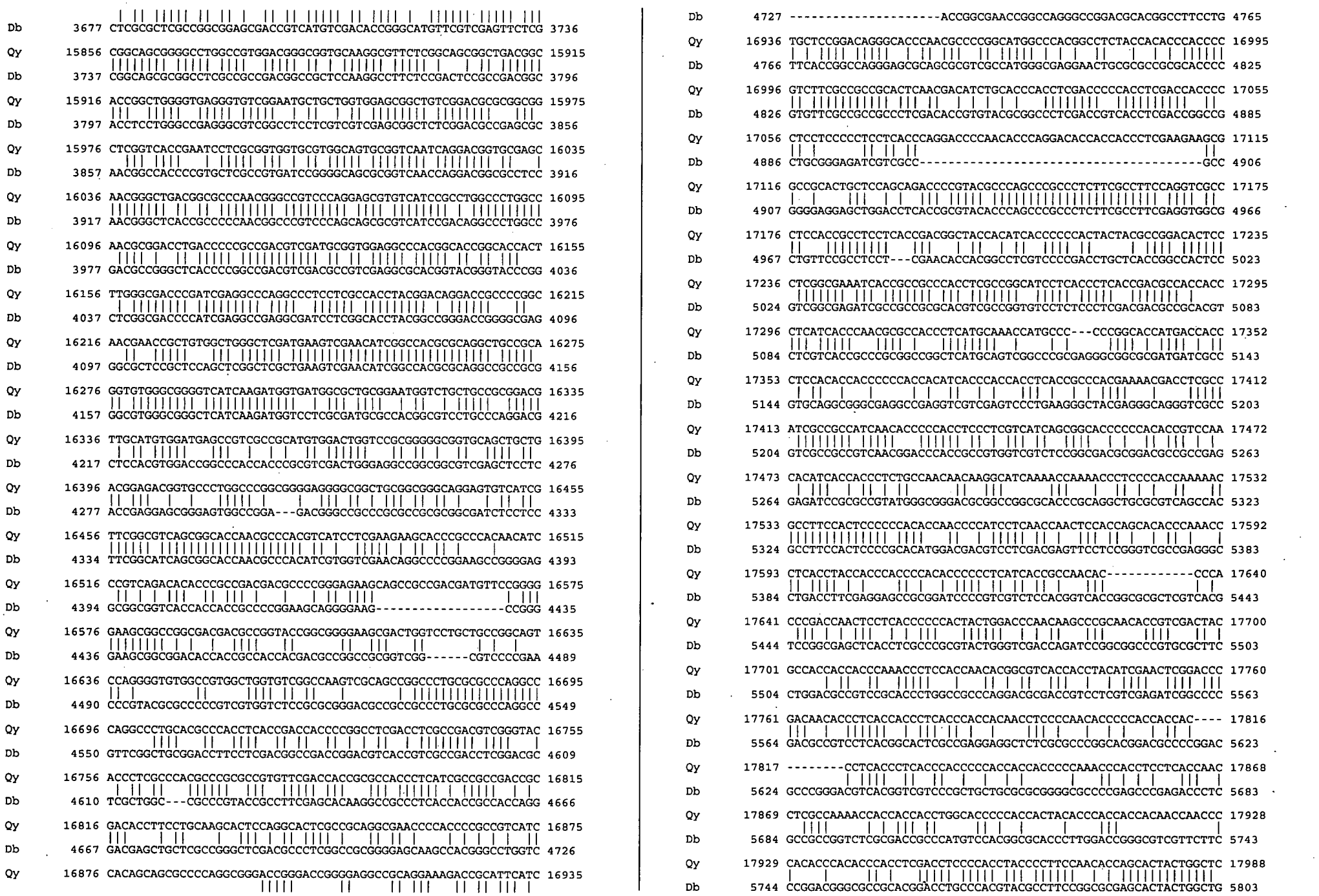
Db        977 GCCGCCGCGCTCGGCGCCGTCCTCGGCTCGGCGCGCCCCGCGGACGAACCCCTGCTCGTC 1036
```

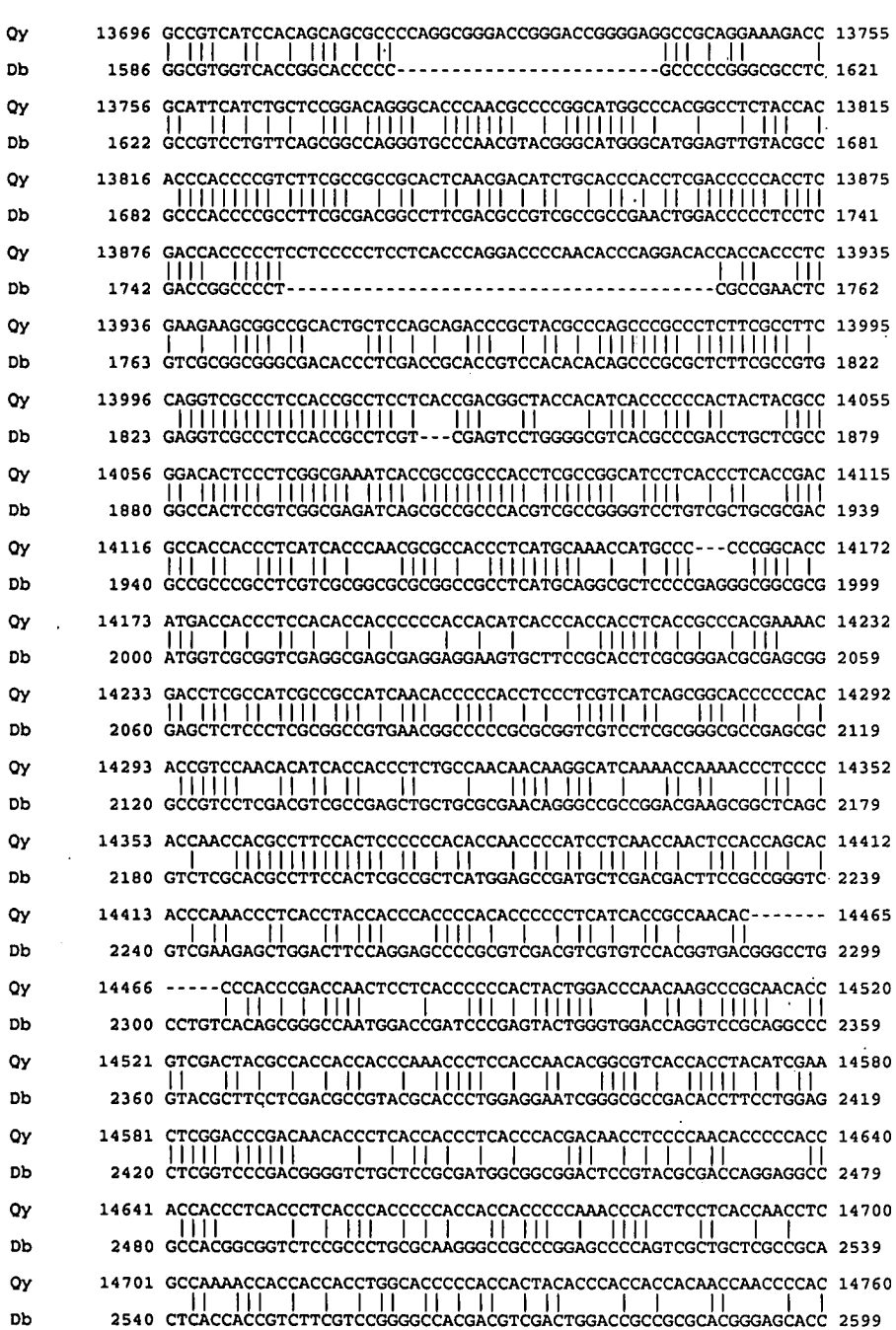

Qy 21694 GGCATGGCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATC 21753
Db 10151 GGGATGGCCCACGACCTGCATGCCGCCCATACCTTCTTCGCGTCCGCCCTCGACGAGGTG 10210
Qy 21754 TGCACCCACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGAC 21813
Db 10211 ACGGACCGTCTCGACCCGCTGCTCGGCCGGCCGCTCGGCGCGCTGCT----------G 10258
Qy 21814 AACGACAACGACAACGAGGACGCGGCCGCACTGCTCCAGCAGACCCCGTACGCCCAGCCC 21873
Db 10259 GACGCCCGACCCGGCTCGCCCGAAGCGGCACTCCTGGACCGGACCGAGTACACCCAGCCG 10318
Qy 21874 GCCCTCTTCGCCTTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACC 21933
Db 10319 GCGCTCTTCGCCGTCGAGGTGGCGCTCCACCGGCTGCTG---GAGCACTGGGGGATGCGC 10375
Qy 21934 CCCCACTACTACGCCGGACACTCCCTCGGCGAAATCACCGCCGCCCCACCTCGCCGGCATC 21993
Db 10376 CCCGACCTGCTGCTGGGGCACTCGGTGGGCGAACTGGCGGCCGCCCACGTCGCGGGTGTG 10435
Qy 21994 CTCACCCTCACCGACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATG 22053
Db 10436 CTCGATCTCGACGACGCCTGCGCGCTGGTGGCCGCCCGCGGCAGGCTGATGCAGCGCCTG 10495
Qy 22054 CCCCC---CGGCACCATGACCACCCTCCACACCACCCCACACCACATCACCCACCACCTC 22110
Db 10496 CCGCCCGGCGGCGCGATGGTCTCCGTGCGGGCCGGCGAGGACGAGGTCCGCGCACTGCTG 10555
Qy 22111 ACCGCCCACGAAAACGACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATC 22170
Db 10556 GCCGGCCGCGAGGACGCCGTCTGCGTCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATC 10615
Qy 22171 AGCGGCACCCCCCACACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAA 22230
Db 10616 TCCGGCGCGGAGGAAGCGGTGGCCGAGGCGGCGGCGCAGCTCGCCGGACGAGGCCGCCGC 10675
Qy 22231 ACCAAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAAC 22290
Db 10676 ACCAGGCGGCTCCGCGTCGCGCACGCCTTCCACTCACCCCTGATGGACGGCATGCTCGCC 10735
Qy 22291 CAACTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACC 22350
Db 10736 GGATTCCGGGAGGTCGCCGCCGGCCTGCGCTACCGGGAACCGGAGCTGACGGTCGTCTCC 10795
Qy 22351 GCCAACA----------CCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAA 22398
Db 10796 ACGGTCACGGGGCGGCCCGCCCGCCCCGGTGAACTCACCGGCCCCGACTACTGGGTGGCC 10855
Qy 22399 CAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTC 22458
Db 10856 CAGGTCCGTGAGCCCGTGCGCTTCGCGGACGCGGTCCGCACGGCACACCGCCTCGGAGCC 10915
Qy 22459 ACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTC 22518
Db 10916 CGCACCTTCCTGGAGACCGGCCCGGACGGCGTGCTGTGCGGCATGGCAGAGGAGTGCCTG 10975
Qy 22519 CCCAACACCCCCACCACCACCCTCACCCTCACCCACCCCCACCACCACCCCCAAACCCAC 22578
Db 10976 GAGGACGACACCGTGGCCCTGCTGCCGGCGATCCACAAGCCCGGCACCGCG--------- 11026
Qy 22579 CTCCTCACCAACCTCGCCAAAACCACCACCACCTGGCACCCCCACCACTACACCCACCAC 22638
Db 11027 ---------- 11026
Qy 22639 CACAACCAACCCCACACCCACACCCACCTCGACCTCCCCACCTACCCCTTCCAACACCAC 22698
Db 11027 ---------- 11026
Qy 22699 CACTACTGGCTCGAACTACCCAGCGCCCAAACCAGCCCCGGTCAAAGGCGTTCTCGCCGC 22758
Db 11027 ----------CCGCACGGTCCGGCGGCTCCCGGCGCGCTGCGGGCGGCCGCC 11068
Qy 22759 TCGGCTCCAGACACCGCCGAGTCGGAGTTCTGGGACGCGGTGAACGAGGAAGACCTCCAG 22818
Db 11069 GCCGCGTACGGCCGGGGCGCCCGGGTGGACTGGGCCGGGATGCACGCCGACGGCCCCGAG 11128
Qy 22819 AGCCTCGCCGAAACCCTCGACATCGACGCCTCTGCTCTGGACACGGTGGTGCCCGCACTC 22878
Db 11129 GGGCCGGCCCGCCGCGTCGAACTGCCCGTCCACGCCTT----------CCGGCAC 11173
Qy 22879 TCCGCCTGGCACCGCCACCAACACGACCAAGCCCGCATCAACACCTGGACCTACCAGGAA 22938
Db 11174 CGCCGCTACTGGCTCGCCCCGGGCCGCGCGGCGGACACCGACGACTGGATGTACCGGATC 11233
Qy 22939 ACCTGGAAACCCCT---CACCCTCCCCACCACCCACCAACCCCACCAAACCTGGCTCATC 22995
Db 11234 GGCTGGGACCGGCTGCCGGCTGTGACCGGCGGGGCCCGGACCGCCGGCCGCTGGCTGGTG 11293
Qy 22996 GCCATCCCCGAAACCCAGACCCACCACCCCCACATCACCAACATCCTCACCAACCTCCAC 23055
Db 11294 ATCCACCCCGACAGCCCGCGCTGCCGGGAGCTGTCCGGCCACGCCGAACGCGCGCTGCGC 11353
Qy 23056 CACCACGGCATCACCCCCATCCCCCTCACTGTCAACCACACCCACACCAACCCCCAACAC 23115
Db 11354 GCCGCGGGCGCGAGCCCCGTACCGCTGCCCGTGGACGCTCCGGCCGCCGACCGGGCGTCC 11413
Qy 23116 CTCCACCACACCCTCCACCA---CACCCGACAACAAGCCCAAAACCACACCACCGGACCC 23172
Db 11414 TTCGCGGCACTGCTGCGCTCCGCCACCGGACCTGACACACGAGGTGACACAGCCGCGCCC 11473
Qy 23173 ATCACCGGCCTGCTCTCCCTCCTCGCCCTCGACGAAACACCCCACCCCCACCACCCCCAC 23232
Db 11474 GTGGCCGGTGTGCTGTCGCTGCTGTCCGAGGAGGATCGGCCCCATCGCCAGCACGCCCCG 11533
Qy 23233 ACACCCACCGGCACCCTCCTCAACCTCACCCTCCCCCAAACCCACACCCAAACCCACCCA 23292
Db 11534 GTACCCGCCGGGGTCCTGGCGACGCTGTCCCTGATGCAGGCTATGGAGGAGGAGGCGGTG 11593
Qy 23293 CCAACCCCCCTCTGGTACGCCACCACCAACGCCACCACCACCCACCCCAACGACCCCCTC 23352
Db 11594 GAGGCTCGCGTGTGGTGCGTCTCCCGCGCCGCGGTCGCCGCCGCCGACCGGGAACGGCCC 11653
Qy 23353 ACACACCCCACCCAAGCCCAAACCTGGGGACTCGCCCGCACCACCCTCCTCGAACACCCC 23412
Db 11654 GTCGGCGCGGGCGCCGCCC---TGTGGGGGCTGGGGCGGGTGGCCGCCCTGGAACGCCCC 11710
Qy 23413 ACCCACACCGCCGGAATCATCGACCTCCCCACCACCCCCACCCCCCACACCCTCCACCAC 23472
Db 11711 ACCCGGTGGGGCGGTCTCGTGGACCTGCCCGCCTCGCCCGGTGCGGCGCACTGGGCGGCC 11770
Qy 23473 CTCACCCAAACCCTCACCCAACCCCACCACCAAACCCAACTCGCCATCCGCACCACCGGC 23532
Db 11771 GCCGTGGAACGGCTCGCCGGTCCCGAGGACCAGATCGCCGTGCGCGCGTCCGGCAGTTGG 11830
Qy 23533 ACCCACACCCGCCGCCTCACCCCCACCACCCTCACCCCCACACACCAACCACCCACCCCC 23592
Db 11831 GGCCGGCGCCTCACCAGGCTGCCGCGCGACGGCGGCGGCCGGACGGCCGCACCCGCGTAC 11890
Qy 23593 ACCCCCCACGGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACC 23652
Db 11891 CGGCCGCGCGGCACGGTGCTCGTCACCGGTGGCACCGGCGCGCTCGGCGGGCATCTCGCC 11950
Qy 23653 CACCACCTCACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCC 23712
Db 11951 CGCTGGCTCGCCGC---GGCGGGCGCCGAACACCTGGCGCTCACCAGCCGCCGGGGCCCG 12007
Qy 23713 CACACCCCCCACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACC 23772
Db 12008 GACGCGCCCGGCGCCGCCGGACTCGAGGCCGAACTCCTCCTCCTGGGCGCCAAGGTGACG 12067
Qy 23773 ATCACCACCTGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCC 23832
Db 12068 TTCGCCGCCTGCGACACCGCCGACCGCGACGGCCTCGCCCGGGTCCTGCGGGCGATACCG 12127
Qy 23833 CCACAACACCCCCTCACCACCGTCATCCACACCGCAGGCGTCAATCTCTTCGCCCCCGTG 23892


Db 7964 CTCGCGCTCTTCGACGCCGCGGTCGCCAGGGACGACGCCCTCCTCGTACCCGCCGGGCTC 8023
Qy 19564 TGGAAGAAATTGGGACCGGTTCTCTCCAGCAAGTCGTCGGTCTTGCTCGAGGACCTTCCC 19623
Db 8024 CGTCCCACCGCCCACCGGGGCACGGACGGACAGCCTCCTGCGCTGTGGCGCGGCCTCGTC 8083
Qy 19624 CAGGCACAGGGAACTGAGGAGGCGCGCAGTACCGTTGAGCAGACGGAGAGCACAAACCTC 19683
Db 8084 CGGGCGCGCCCGCGCCGTGCCGCGCGGACGGCCGCCGAGGCGGCGGACACGACCGGCGGC 8143
Qy 19684 CGGCAACTC---CTCATGGGTCGGTCACGTTCCGAGCAGGAAGAAGAGCTGCTCAGCCTC 19740
Db 8144 TGGCTGAGCGGGCTCGCCGCACAGTCCCCCGAGGAGCGGCGCAGCACAGCCGTCACGCTC 8203
Qy 19741 GTCCGCATCCACTCCGCGGCAGTGCTCGGGCGCGACGACTCCGAGGCCATCCCGCCCGGT 19800
Db 8204 GTGACGGGTGTCGTCGCGGACGTCCTCGGGCACGCCGACTCCGCCGCGGTCGGGGCGGAG 8263
Qy 19801 CGGCTGTTCAGGGATCTAGGGTTCGACTCGCTTGCGGCGGTGGAGCTTCGCAACCACCTC 19860
Db 8264 CGGTCCTTCAAGGACCTCGGCTTCGACTCCCTGGCCGGGGTGGAGCTCCGCAACCGGCTG 8323
Qy 19861 GCAGCACAGACGGAGCTGGCTCTGCCGACGACTCTCGTCTTCGATTACCCCAGCCCCACC 19920
Db 8324 AACGCCGCCACCGGCCTGCGGCTCCCCGCGACCACGGTCTTCGACCATCCCTCGCCGGCC 8383
Qy 19921 AAGCTCGCCCAATTTCTGCTCTCCGAGATCGCGGAGTTCCAGCCCGACAACTCAACTCCG 19980
Db 8384 GCGCTCGCGTCCCATCTCCTCGCCCAGGTGCCCGGGTTGAAGGAGGGGACGGCGGCGACC 8443
Qy 19981 CTTCC----------GCGACCCCGGGCAGAGCTCGATGAGCCG 20013
Db 8444 GCGACCGTCGTGGCCGAGCGGGGCGCTTCCTTCGGTGACCGTGCGACCGACGACGATCCG 8503
Qy 20014 ATCGCCATCGTTGGCATGGCCTGTCGCTTCCCCGGCGGAGTGACCTCGGCGGACGACTTC 20073
Db 8504 ATCGCGATCGTGGGCATGGCATGCCGCTATCCGGGTGGTGTGTCGTCGCCGGAGGACCTG 8563
Qy 20074 TGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGCGGCTGG 20133
Db 8564 TGGCGGCTGGTGGCCGAGGGGACGGACGCGATCAGCGAGTTCCCCGTCAACCGCGGCTGG 8623
Qy 20134 GACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACCCGAAAC 20193
Db 8624 GACCTGGAGAGCCTCTACGACCCGGATCCCGAGTCGAAGGGCACCACGTACTGCCGGGAG 8683
Qy 20194 GGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGCCCCCGC 20253
Db 8684 GGCGGGTTCCTGGAAGGCGCCGGTGACTTCGACGCCGCCTTCTTCGGCATCTCGCCGCGC 8743
Qy 20254 GAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAAACCATC 20313
Db 8744 GAGGCCCTGGTGATGGACCCGCAGCAGCGGCTGCTGCTGGAGGTGTCCTGGGAGGCGCTG 8803
Qy 20314 GAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTCACCGGC 20373
Db 8804 GAACGCGCGGGCATCGACCCGTCCTCGCTGCGCGGCAGCCGCGGTGGTGTCTACGTGGGC 8863
Qy 20374 ACCAACGGACAGGACCACGCGGCACACATCCGTCAGGCCCCGAGCGGTACCGAGGGATTC 20433
Db 8864 GCCGCGCACGGCTCGTACGCCTCCGATCCCCGGCTGGTGCCCGAGGGCTCGGAGGGCTAT 8923
Qy 20434 GTCCTGACCGGGGCAGCCACCAGCATCGCCTCCGGCCGAATCTCCTACATCCTCGGGTTG 20493
Db 8924 CTGCTGACCGGCAGCGCCGACGCGGTGATGTCCGGCCGCATCTCCTACGCGCTCGGTCTC 8983
Qy 20494 GAAGGGCCTGCGGTCACCCTCGACACAGCGTGTTCCTCCTCGCTCGTCGCCCTGCACCTC 20553
Db 8984 GAAGGACCGTCCATGACGGTGGAGACGGCCTGCTCCTCCTCGCTGGTGGCGCTGCATCTG 9043
Qy 20554 GCCTGCCAGTCCCTCAGGTCCGGTGAATGCACCATGGCCTTGGCCGGCGGGGCCACGGTC 20613
Db 9044 GCGGTACGGGCGCTGCGGCACGGCGAGTGCGGGCTCGCGCTGGCGGGCGGGGTGGCGGTG 9103
Qy 20614 ATGACCACCCCGATCACCTTCACCGAATTCGCCCGCCAACGCGGACTCGCCCCCGACGGG 20673
Db 9104 ATGGCCGATCCGGCGGCGTTCGTGGAGTTCTCCCGGCAGAAGGGGCTGGCCGCCGACGGC 9163
Qy 20674 CGTTGCAAGGCGTTCTCGGCGGCGGCTGACGGTACCGGCTGGGGTGAGGGTGTGGGGATG 20733
Db 9164 CGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGGCTGGGCCGAGGGCGTCGGCGTG 9223
Qy 20734 CTGCTGGTGGAGCGGCTCTCCGACGCCCGCCGCAACGGTCACCGTGTCCTGGCCGTGGTG 20793
Db 9224 CTCGTCCTGGAGCGGCTGTCGGACGCGCGCCGCGCGGGGCACACGGTCCTCGGCCTGGTC 9283
Qy 20794 CGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAACGGGCCC 20853
Db 9284 ACCGGCACCGCGGTCAACCAGGACGGTGCCTCCAACGGGCTGACCGCGCCCAACGGCCCA 9343
Qy 20854 TCCCAGCAGCGCGTCATCCGCCAGGCCCTCGCCAACGCGGACCTGACCCCCGCCGACGTC 20913
Db 9344 GCCCAGCAACGCGTCATCGCCGAGGCGCTCGCCGACGCCGGGCTGTCCCCGGAGGACGTG 9403
Qy 20914 GATGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCCCAGGCC 20973
Db 9404 GACGCGGTCGAGGCGCACGGCACCGGCACCCGGCTCGGCGACCCCATCGAGGCCGGGGCG 9463
Qy 20974 ATCCTCGCGACCTACGGACAGGACCGTCCCGGCAACGGGCCGTTGTGGCTGGGCTCCGTC 21033
Db 9464 CTGCTCGCCGCCTCCGGACGGAACCGTTCCGGCGACCACCCGCTGTGGCTCGGCTCGCTG 9523
Qy 21034 AAGTCCAACGTCGGACACACACAGGCCGCGGCGGGCGTGGCCGGAGTGATCAAGATGGTG 21093
Db 9524 AAGTCCAACATCGGGCATGCCCAGGCCGCCGCCGGTGTCGGCGGCGTCATCAAGATGCTC 9583
Qy 21094 ATGGCCCTCCGCCACCGGACACTCCCACCGACTCTCCACGCGGATGAGCCGTCGCCGCAT 21153
Db 9584 CAGGCGCTGCGGCACGGCTTGCTGCCCCGCACCCTCCACGCCGACGAGCCGACCCCGCAT 9643
Qy 21154 GTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCCGGCGGG 21213
Db 9644 GCCGACTGGAGCTCCGGCCGGGTACGGCTGCTCACCTCCGAGGTGCCGTGGC---AGCGG 9700
Qy 21214 GAGGGGCGGCCGCGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAACGCCCAC 21273
Db 9701 ACCGGCCGGCCCCGGCGGACCGGGGTGTCCGCCTTCGGCGTCGGCGGCACCAATGCCCAT 9760
Qy 21274 GTCATCCTCGAAGAAGCACCCGCCGACGACGTTCCGGGGGGGACCACCCGCCGACGAGGAT 21333
Db 9761 GTCGTCCTCGAAGAGGCACCCGCC---------CCGCCCGCGCCGGAACCGGCCGGGGAG 9811
Qy 21334 GCCGGTAGTGGCGAGGAGGCTGCTGCCGGCAGTCCTGGGGTGTGGCCGTGGCTGGTGTCG 21393
Db 9812 GCCCCCGGCGGCTCCCGCGCCGCAGAAGGGGCGGAAGGGCCCCTGGCCTGGGTGGTCTCC 9871
Qy 21394 GCCAAGTCGCAGCCGGCCCTGCGCGCCCAGGCCCAGGCCCTGCACGCCCACCTCACCGAC 21453
Db 9872 GGACGCGACGAGCCGGCCCTGCGGTCCCAGGCCCGGCGGCTCCGCGACCACCTCTCCCGC 9931
Qy 21454 CACCCCGGCCTCGACCTCGCCGACGTCGGATACACCCTCGCCCACGCCCGCGCCGTGTTC 21513
Db 9932 ACCCCCGGGGCCCGCCCGCGTGACATCGCCTTCTCCCTCGCCGCCACGCGCGCAGCCTTT 9991
Qy 21514 GACCACCGCGCCACCCTCATCGCCGCCGACCGCGACACCTTCCTGCAAGCACTCCAGGCA 21573
Db 9992 GACCACCGCGCCGTGCTGATCGGCTCGGACGGGGCCGAACTCGCCGCCGCCCTGGACGCG 10051
Qy 21574 CTCGCCGCAGGCGAACCCCACCCCGCCGTCATCCACAGCAGCGCCCCAGGCGGGACCGGG 21633
Db 10052 TTGGCCGAAGG--------ACGCGACGGTCCGGCGGTGGTGCGCGGAGTCCGCGACCGGG 10103
Qy 21634 ACCGGGGAGGCCGCAGGAAAGACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCC 21693
Db 10104 AC----------GGCAGGATGGCCTTCCTCTTCACCGGGCAGGGCAGCCAGCGCGCC 10150








Qy 13696 GCCGTCATCCACAGCAGCGCCCCAGGCGGGACCGGGACCGGGGAGGCCGCAGGAAAGACC 13755
Db 1586 GGCGTGGTCACCGGCACCCCC----------GCCCCCGGGCGCCTC 1621
Qy 13756 GCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCAC 13815
Db 1622 GCCGTCCTGTTCAGCGGCCAGGGTGCCCAACGTACGGGCATGGGCATGGAGTTGTACGCC 1681
Qy 13816 ACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTC 13875
Db 1682 GCCCACCCCGCCTTCGCGACGGCCTTCGACGCCGTCGCCGCCGAACTGGACCCCCTCCTC 1741
Qy 13876 GACCACCCCCTCCTCCCCCTCCTCACCCAGGACCCCAACACCCAGGACACCACCACCCTC 13935
Db 1742 GACCGGCCCCT----------CGCCGAACTC 1762
Qy 13936 GAAGAAGCGGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTC 13995
Db 1763 GTCGCGGCGGGCGACACCCTCGACCGCACCGTCCACACACAGCCCGCGCTCTTCGCCGTG 1822
Qy 13996 CAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCC 14055
Db 1823 GAGGTCGCCCTCCACCGCCTCGT---CGAGTCCTGGGGCGTCACGCCCGACCTGCTCGCC 1879
Qy 14056 GGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGAC 14115
Db 1880 GGCCACTCCGTCGGCGAGATCAGCGCCGCCCACGTCGCCGGGGTCCTGTCGCTGCGCGAC 1939
Qy 14116 GCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCC---CCCGGCACC 14172
Db 1940 GCCGCCCGCCTCGTCGCGGCGCGCGGCCGCCTCATGCAGGCGCTCCCCGAGGGCGGCGCG 1999
Qy 14173 ATGACCACCCTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAAC 14232
Db 2000 ATGGTCGCGGTCGAGGCGAGCGAGGAGGAAGTGCTTCCGCACCTCGCGGGACGCGAGCGG 2059
Qy 14233 GACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCAC 14292
Db 2060 GAGCTCTCCCTCGCGGCCGTGAACGGCCCCCGCGCGGTCGTCCTCGCGGGCGCCGAGCGC 2119
Qy 14293 ACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCC 14352
Db 2120 GCCGTCCTCGACGTCGCCGAGCTGCTGCGCGAACAGGGCCGCCGGACGAAGCGGCTCAGC 2179
Qy 14353 ACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCAC 14412
Db 2180 GTCTCGCACGCCTTCCACTCGCCGCTCATGGAGCCGATGCTCGACGACTTCCGCCGGGTC 2239
Qy 14413 ACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACAC------- 14465
Db 2240 GTCGAAGAGCTGGACTTCCAGGAGCCCCGCGTCGACGTCGTGTCCACGGTGACGGGCCTG 2299
Qy 14466 -----CCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACC 14520
Db 2300 CCTGTCACAGCGGGCCAATGGACCGATCCCGAGTACTGGGTGGACCAGGTCCGCAGGCCC 2359
Qy 14521 GTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAA 14580
Db 2360 GTACGCTTCCTCGACGCCGTACGCACCCTGGAGGAATCGGGCGCCGACACCTTCCTGGAG 2419
Qy 14581 CTCGGACCCGACAACACCCTCACCACCCTCACCCACGACAACCTCCCCAACACCCCCACC 14640
Db 2420 CTCGGTCCCGACGGGGTCTGCTCCGCGATGGCGGCGGACTCCGTACGCGACCAGGAGGCC 2479
Qy 14641 ACCACCCTCACCCTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTCACCAACCTC 14700
Db 2480 GCCACGGCGGTCTCCGCCCTGCGCAAGGGCCGCCCGGAGCCCCAGTCGCTGCTCGCCGCA 2539
Qy 14701 GCCAAAACCACCACCACCTGGCACCCCCACCACTACACCCACCACCACAACCAACCCCAC 14760
Db 2540 CTCACCACCGTCTTCGTCCGGGGCCACGACGTCGACTGGACCGCCGCGCACGGGAGCACC 2599
Qy 14761 ACCCACACCCACCTCGACCTCCCCACCTACCCCTTCCAACACCACCACTACTGGCTCCA- 14819
Db 2600 GGCACGGTCAGGGTGCCCCTGCCGACCTACGCCTTCCAGCGCGAACGCCACTGGTTCGAC 2659
Qy 14820 ACCACCCGGCAAGCCGAGCGACCCGTCACCGAGCGAAGGCCGTGAGCAAGCCACGACCCC 14879
Db 2660 GGCGCCGCGCGAACGGCGGCGCCGCTCACGGCGGGCCGATCGGGCACCGGTGCGGGCACC 2719
Qy 14880 ATCAACCCCGCTGCGTGATGTCCTCGTGGGCAAGTCTCCGCAGGAGCGAGACGAAGAGCT 14939
Db 2720 GGCCCGGCCGCGGGTGTGACGTCGGGCGAGGGCGAGGGCGAGGGCGAGGGCGCGGGTGCG 2779
Qy 14940 GTTGCGCCTG----------GTGCGCACCCATGCGGCC 14967
Db 2780 GGTGGCGGTGATCGGCCGGCTCGCCACGAGACGACCGAGCGCGTGCGCGCACACGTCGCC 2839
Qy 14968 GCTGTGCTGGGCCATGCCACTCCCGAAGTGATCGTTCCGAACAAGGCCTTCAAAGAGCTG 15027
Db 2840 GCCGTCCTCGAGTACGACGACCCGACCCGCGTCGAACTCGGCCTCACCTTCAAGGAGCTG 2899
Qy 15028 GGTTTTGATTCTCTCGCCGCAATTCAGCTTCGTAATCGACTGCTTGCTGACGTTGACCTG 15087
Db 2900 GGCTTCGACTCCCTCATGTCCGTCGAGCTGCGGAACGCGCTCGTCGACGACACGGGACTG 2959
Qy 15088 CCGCTTCCGGCCACGCTGATCTTCGATTACCCCACTCCGATGGCGCTT---------- 15135
Db 2960 CGCCTGCCCAGCGGACTGCTCTTCGACCACCCGACGCCGCGCGCCCTCGCCGCCCACCTG 3019
Qy 15136 TGCCAGTTCCTCCGGGCGGCGATCGTCGGAGCGGACACAGGCACGACCACTCGTCTGCCG 15195
Db 3020 GGCGACCTGCTCACCGGCGGCAGCGGCGAGACCGGATCGGCCGACGGGATACCGCCCGCG 3079
Qy 15196 CTAACTGCGGTCCCCGCCGACGAGCCGATCGCCATCGTCGGCATGGCCTGTCGGTACCCC 15255
Db 3080 ACCCCGGCGGACACCACCGCCGAGCCCATCGCGATCATCGGCATGGCCTGCCGCTACCCC 3139
Qy 15256 GGTGATGTACGGACGGTCGATGATCTCTGGCAGGTGGTCAGTGGTGGCCATGACGCGATC 15315
Db 3140 GGCGGCGTCACCTCCCCCGAGGACCTGTGGCGGCTCGTCGCCGAGGGGCGCGACGCCGTC 3199
Qy 15316 GGCGGATTCCCGACGAACCGTGGGTGGGACCTCGACACGCTGTACAACCCGGACCCGGAC 15375
Db 3200 TCGGGGCTGCCCACCGACCGCGGCTGGGACGAGGAC---CTCTTCGACGCCGACCCCGAC 3256
Qy 15376 CACCACGGAACCAGCTACACCCGGAGCGGCGGATTCCTTTACGACGCAGGCAATTTCGAT 15435
Db 3257 CGCAGCGGCAAGAGCTCGGTCCGCGAGGGCGGATTCCTGCACGACGCCGCCCTGTTCGAC 3316
Qy 15436 CCCGACTTCTTCGGTATCAGTCCGCGTGAGGCACTGGCGATGGACCCGCAGCAGCGGCTG 15495
Db 3317 GCCGGCTTCTTCGGGATATCGCCCCGCGAGGCCCTCGGCATGGACCCGCAGCAGCGGCTG 3376
Qy 15496 CTGCTGGAAACAGCGTGGGAGAGCATCGAACACGCCTGCATCAACCCCGACAGCCTCCGT 15555
Db 3377 CTCCTGGAGACGGCATGGGAGGCCGTGGAGCGCGCAGGGCTCGACCCCGAAGGCCTCAAG 3436
Qy 15556 GGCACACCAACCGGCGTCTTCGCCGGGCTGACCTACCACGACTACGCCGCGCGCTTTCCC 15615
Db 3437 GGCAGCCGGACGGCCGTCTTCGTCGGCGCCACCGCCCTGGACTACGGCCCGCGCATGCAC 3496
Qy 15616 ACAGCTCCGGCAGGGTTCGAGGGGTATCTCGGGCACGGAAGCGCAGGCAGTATCGCCTCG 15675
Db 3497 GACGGCGCCGAGGGCGTCGAGGGCCACCTCCTGACCGGGACCACGCCCAGCGTGATGTCG 3556
Qy 15676 GGTCGTGTCGCCTACGCTCTCGGCCTGGAAGGTCCGGCCCTCACAGTCGACACTGCCTGC 15735
Db 3557 GGCCGCATCGCCTACCAGCTCGGCCTCACCGGTCCTGCGGTCACCGTCGACACGGCCTGC 3616
Qy 15736 TCTTCGTCCCTGGTCGCTCTGCACCTGGCCTGTCAGGCGCTGCGGTCCGGCGAGTGTTCC 15795
Db 3617 TCGTCCTCGCTCGTCGCGCTGCACCTGGCCGTCCGTTCGCTGCGGCAGGGCGAGTCGAGC 3676
Qy 15796 ATGGCCCTCGCGGGTGGCGTCACGGTGATGTCAACCCCGGCCGGGTTCGTGGAGTTTTCG 15855

```
Db      12698 GCTGCGGCGGACC---TCACCGCCCTGCCGCCCGCCGAACTCCGCACCGCGCTGCGCGAG 12754

Qy      24490 CTGATTCGGACAGGCATCTGCACCGTTCTCGGCCTTCGTAATCCGGAAGGCATCGAGGAC 24549

Db      12755 CTGGTGCGAGCCCGGACCGCCGCGGCGCTCGGCCTCGACGACCCGGCCGAGGTCGCCGAG 12814

Qy      24550 CAACGAGCCTTCCGCGACCTCGGCTTCGACTCGCTGACGTCGGCTCAGTTCAGCAAGGAA 24609

Db      12815 GGCGAACGGTTCCCCGCCATGGGCTTCGACTCCCTGGCCACCGTACGGCTGCGCCGCGGA 12874

Qy      24610 CTCGCCAAGGAAACCGGACTGCCACTCCCCCCGTCCCTGGTCTTCGACTATCCCACCCCG 24669

Db      12875 CTCGCCTCGGCCACGGGCCTCGACCTGCCCCCCGATCTGCTCTTCGACCGGGACACCCCG 12934

Qy      24670 CAGGAATGTGCTGCCCATCTGCGCACACAACTCGTCGACCTAGACGACGAAGAGGACGCG 24729

Db      12935 GCCGCGCTCGCCGCCCACCTGGCCGAACTGCTCGCCACCGCACGGGACCACGGACCCGGC 12994

Qy      24730 GCACTGTCGAATGCTCTCCCGC 24751

Db      12995 GGCCCCGGGACCGGTGCCGCGC 13016


RESULT 13
US-09-836-821-1
; Sequence 1, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/836,821
;  CURRENT FILING DATE:  2001-04-17
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-836-821-1

  Query Match             7.1%;  Score 2188.4;  DB 9;  Length 15872;
  Best Local Similarity   52.5%;  Pred. No. 0;
  Matches 6973;  Conservative    0;  Mismatches 5061;  Indels 1248;  Gaps   50;

Qy      12256 CGCAGCGGAGGGTTTCTCTATGACGCGGGCGACTTCGACGCCGCGTTCTTCGAGTTGTCA 12315

Db        197 CGCCGGGGCGGCTTCCTCGACTCCCTCGACCTCTTCGACGCGGCCTTCTTCGGAATCTCG 256

Qy      12316 CCGCGTGAGGCGCTGGCAATGGACCCGCAGCAGCGCCTGCTGCTCGAAACCACTTGGGAA 12375

Db        257 CCCCGTGAGGCCGCCGCCATGGACCCGCAGCAGCGACTGGTCCTCGAACTCGCCTGGGAG 316

Qy      12376 ACGTTCGAACAGGGCGGAATCGACCCGAGGTCCATGCGCGGAAGCCGGACCGGGGTTTTC 12435

Db        317 GCGCTGGAGGACGCCGGAATCGTCCCCGGCACCCTCGCCGGAAGCCGCACCGCCGTCTTC 376

Qy      12436 GTGGGGATCAATCCGGAGGACTACACCACCGGATACACACATCAGCCCTCAAACGCAGTC 12495

Db        377 GTCGGCACCCTGCGGGACGACTACACGAGCCTCCTCTACCAGCACGGCGAGCAGGCCATC 436

Qy      12496 GAGGGCTACCTGCTCACTGGCAGCGCGGCAAGCATTGCGTCAGGCCGTATCTCCTACAAC 12555

Db        437 ACCCAGCACACCATGGCGGGCGTGAACCGGGGCGTCATCGCCAACCGCGTCTCGTACCAC 496

Qy      12556 TTCGGGCTCGAAGGCCCTGCGATCACTATCGACACCGCGTGTTCCTCCTCGCTCGTCGCC 12615

Db        497 CTCGGCCTGCAGGGCCCGAGCCTCACCGTCGACGCCGCGCAGTCGTCCTCGCTCGTCGCC 556

Qy      12616 CTGCATCTGGCCTGCCAAGCGCTCCGGTCCGGTGAATGCACCATGGCGCTCGCAGGCGGC 12675

Db        557 GTGCACCTGGCCTGCGAGTCCCTGCGCGCCGGGGAGTCCACGACGGCGCTCGTCGCCGGC 616

Qy      12676 GCCTCCGTCATGGCCACTCCCTTCGTCTTCACCGAGTTCTCTCGCCAGCGGGGCCTGGCC 12735

Db        617 GTGAACCTCAACATCCTCGCGGAGAGCGCCGTGACGGAGGAGCGCTTCGGTGGACTCTCC 676

Qy      12736 GCAGACGGCCGGTGCAAGGCGTTTTCGGCGGCGGCGGACGGGACCGGCTGGTCCGAGGGT 12795

Db        677 CCGGACGGCACCGCCTACACCTTCGACGCGCGGGCCAACGGATTCGTCCGGGGCGAGGGC 736

Qy      12796 GTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCGCCGCAACGGTCACCGTGTCCTG 12855

Db        737 GGCGGAGTCGTCGTACTCAAGCCGCTCTCCCGCGCCCTCGCCGACGGCGACCGTGTCCAC 796

Qy      12856 GCCGTCGTCCGCGGCAGCGCCGTCAACCAGGACGGCGCAAGCAACGGCCTGACCGCACCC 12915

Db        797 GGCGTCATCCGCGCCAGCGCCGTCAACAACGACGGAGCCACCCCGGGTCTCACCGTGCCC 856

Qy      12916 AACGGTCGTTCACAAGTCAAGGTCATCCGCCAGGCTTTGGCCAACGCACACCTCTCCCCT 12975

Db        857 AGCAGGGCCGCCCAGGAGAAGGTGCTGCGCGAGGCGTACCGGAAGGCGGCCCTGGACCCG 916

Qy      12976 GCCGATGTCGATGCGGTGGAGGCCCACGGCACGGGGACCACCCTGGGCGACCCGATCGAG 13035

Db        917 TCCGCCGTCCAGTACGTCGAACTCCACGGCACCGGAACCCCCGTCGGCGACCCCATCGAG 976

Qy      13036 GCTCAAGCCCTCGTCGAAGCCTACGGTCAGGACCGCCCCAACGGCCGCCCCCTCTGGCTC 13095

Db        977 GCCGCCGCGCTCGGCGCCGTCCTCGGCTCGGCGCGCCCCGCGGACGAACCCCTGCTCGTC 1036

Qy      13096 GGAACCCTCAAGTCCAACATCGGGCACTCCATGGCCGCTGCGGGTGTGGGCGGGGTCATC 13155

Db       1037 GGCTCGGCCAAGACGAACGTCGGGCACCTCGAAGGCGCCGCCGGCATCGTCGGCCTCATC 1096

Qy      13156 AAGATGGTGATGGCGCTGCGGAATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCG 13215

Db       1097 AAGACGCTCCTCGCGCTCGGCCGGCGCCGGATCCCGGCGAGCCTCAACTTCCGTACGCCC 1156

Qy      13216 TCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGG 13275

Db       1157 CACCCGGACATCCCGCTCGACACCCTCGGGCTCGACGTGCCCGACGGCCTGCGGGAGTGG 1216

Qy      13276 CCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 13335

Db       1217 CCGCACCCGGAC---CGCGAACTCCTCGCCGGCGTCAGCTCGTTCGGCATGGGCGGCACC 1273

Qy      13336 AACGCCCACGTCATCCTCGAGGAAGCACCCGCCCACAACATCCCGTCAGACACACCCGCC 13395

Db       1274 AACGCCCACGTCGTCCTCAGCGAAGGCCCCGCCCAG---------------------- 1309

Qy      13396 GACGACGTCCCGGGAGAATCAGCCGCCGACGAGGATGCCGGTAGTGGCGATGAGGCTGCT 13455

Db       1310 ----------------------GGCGGCGAGCAGCCCGGCATCGATGAGGAGACCCCC 1345

Qy      13456 GCCGGCAGTCCAGGGGTGTGGCCGTGGCTGGTGTCGGCCAAGTCGCAGCCGGCCCTGCGC 13515

Db       1346 GTCGACAGCGGGGCCGCACTGCCCTTCGTCGTCACCGGCCGCGGCGGCGAGGCCCTGCGC 1405

Qy      13516 GCCCAGGCCCAGGCCCTGCACGCCCACCTCACCGACCACCCCGGCCTCGACCTCGCCGAC 13575

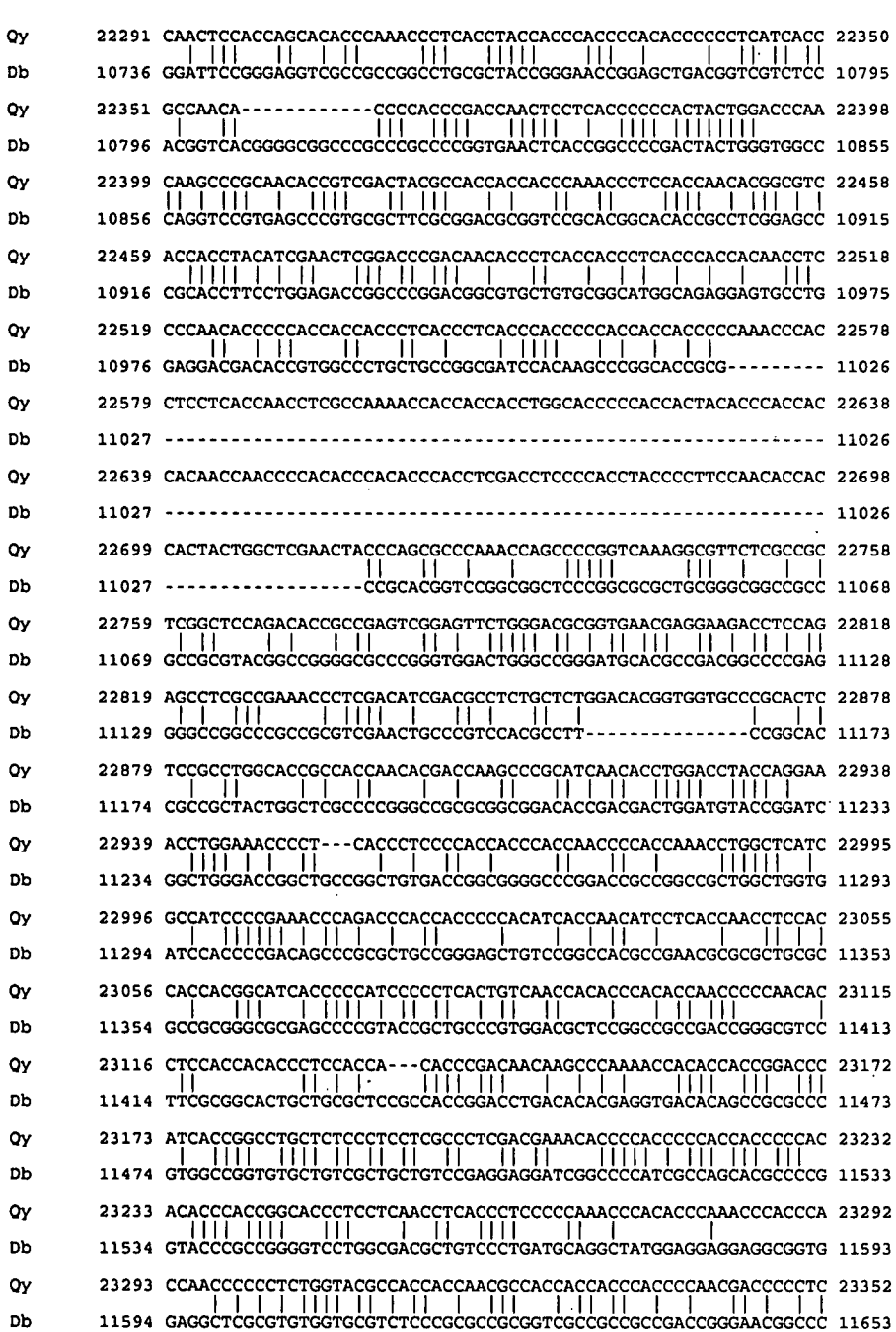
Db       1406 GCCCAGGCCCGGCGCCTGCACGAGGCCGTCGAAGCGGACCCGGAGCTCGCGCCCGCCGCA 1465

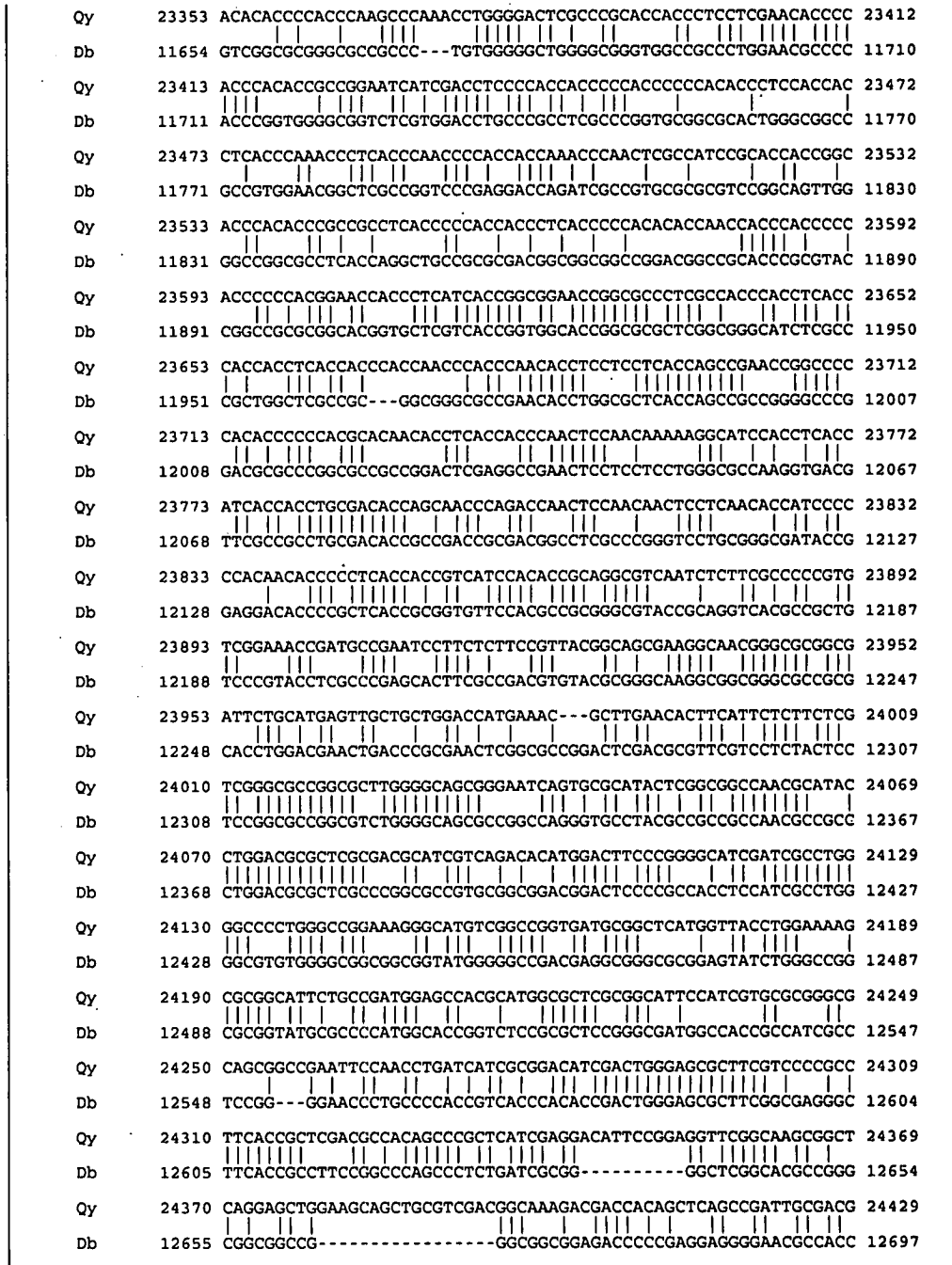
Qy      13576 GTCGGGTACACCCTCGCCCACGCCCGCGCCGTGTTCGACCACCGCGCCACCCTCATCGCC 13635

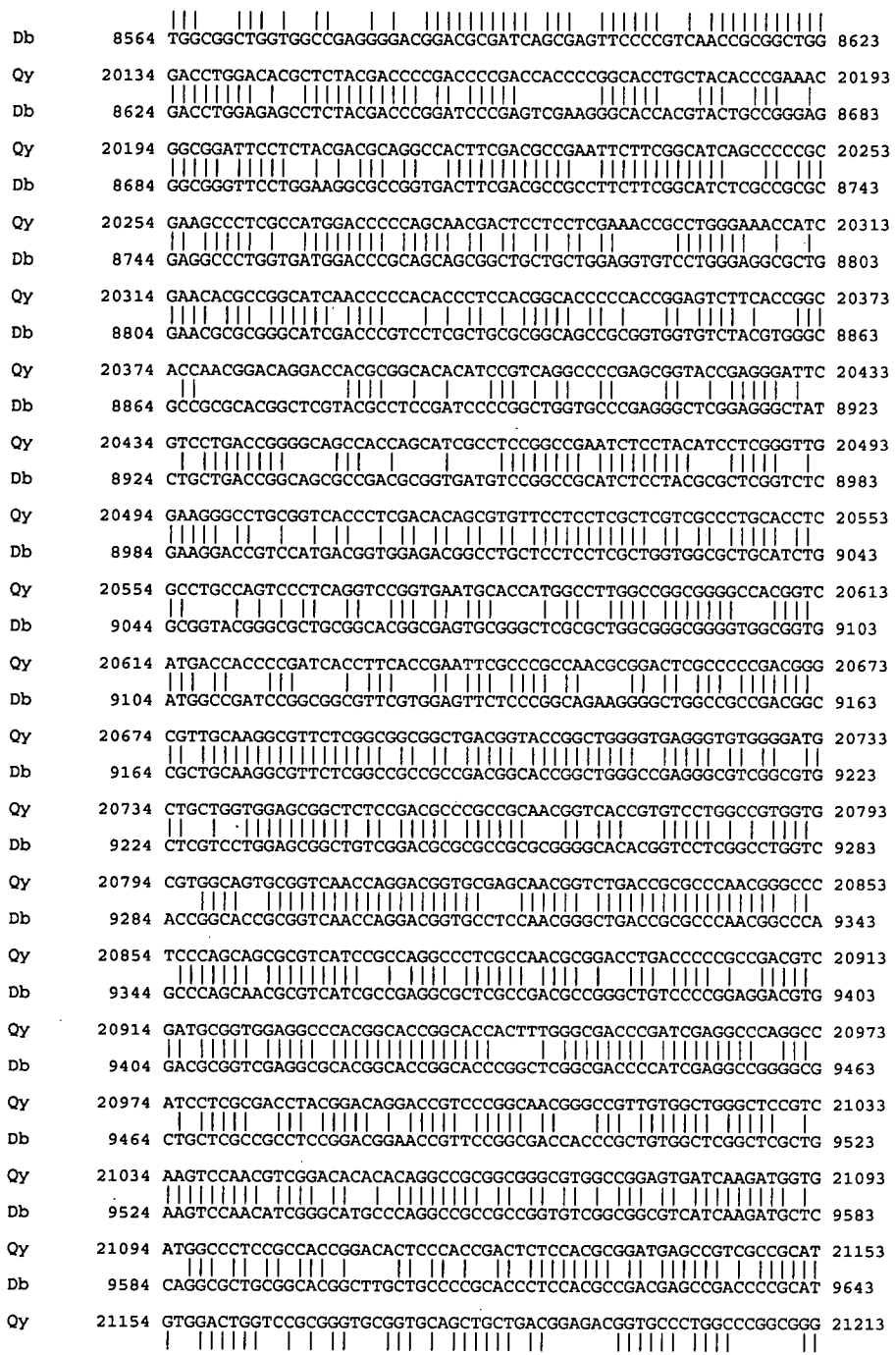
Db       1466 CTCGCCCGGTCGCTGGTCACCACCCGTACGGTCTTCACGCACCGGTCGGTCGTCCTCGCC 1525

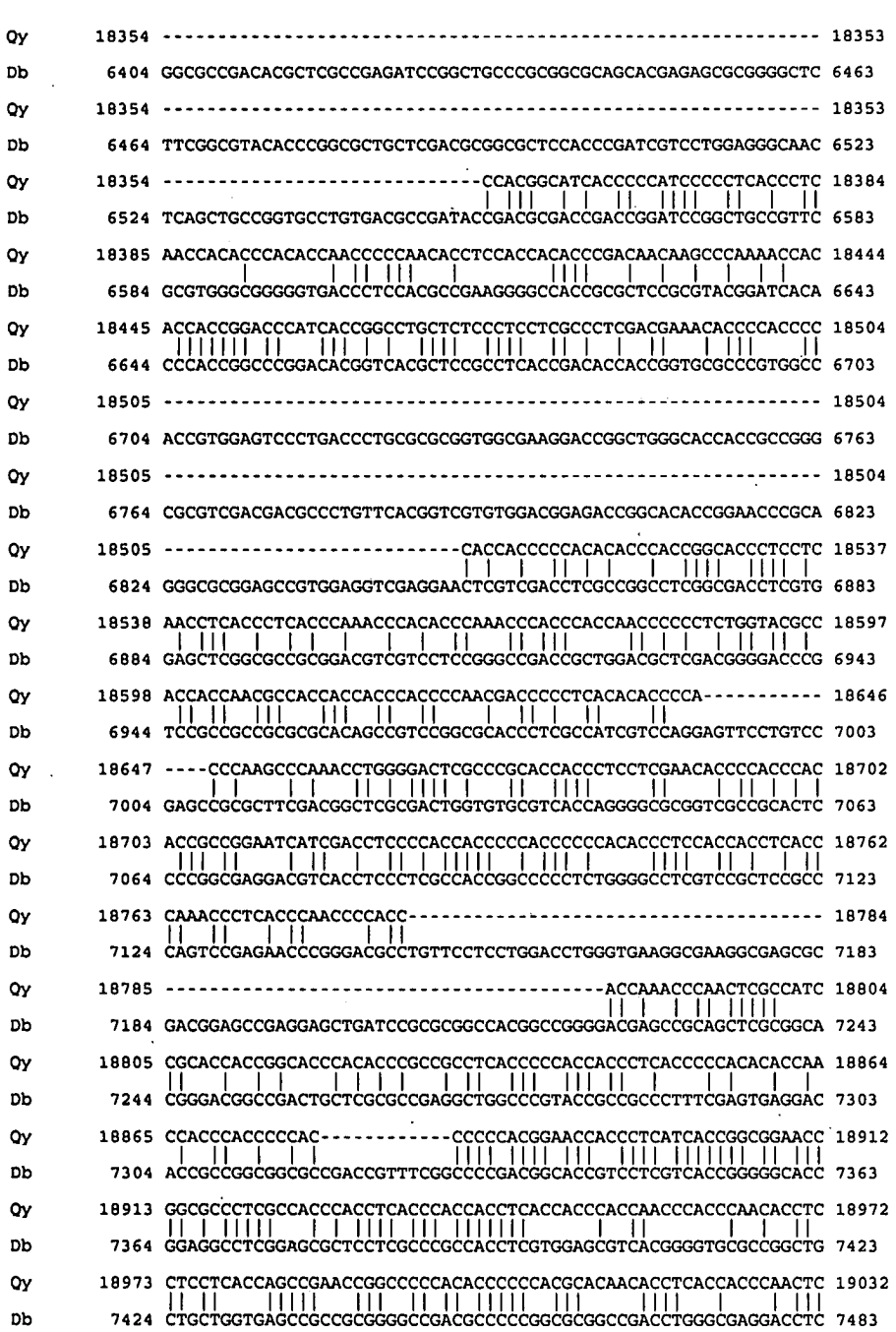
Qy      13636 GCCGACCGCGACACCTTCCTGCAAGCACTCCAGGCACTCGCCGCAGGCGAACCCCACCCC 13695

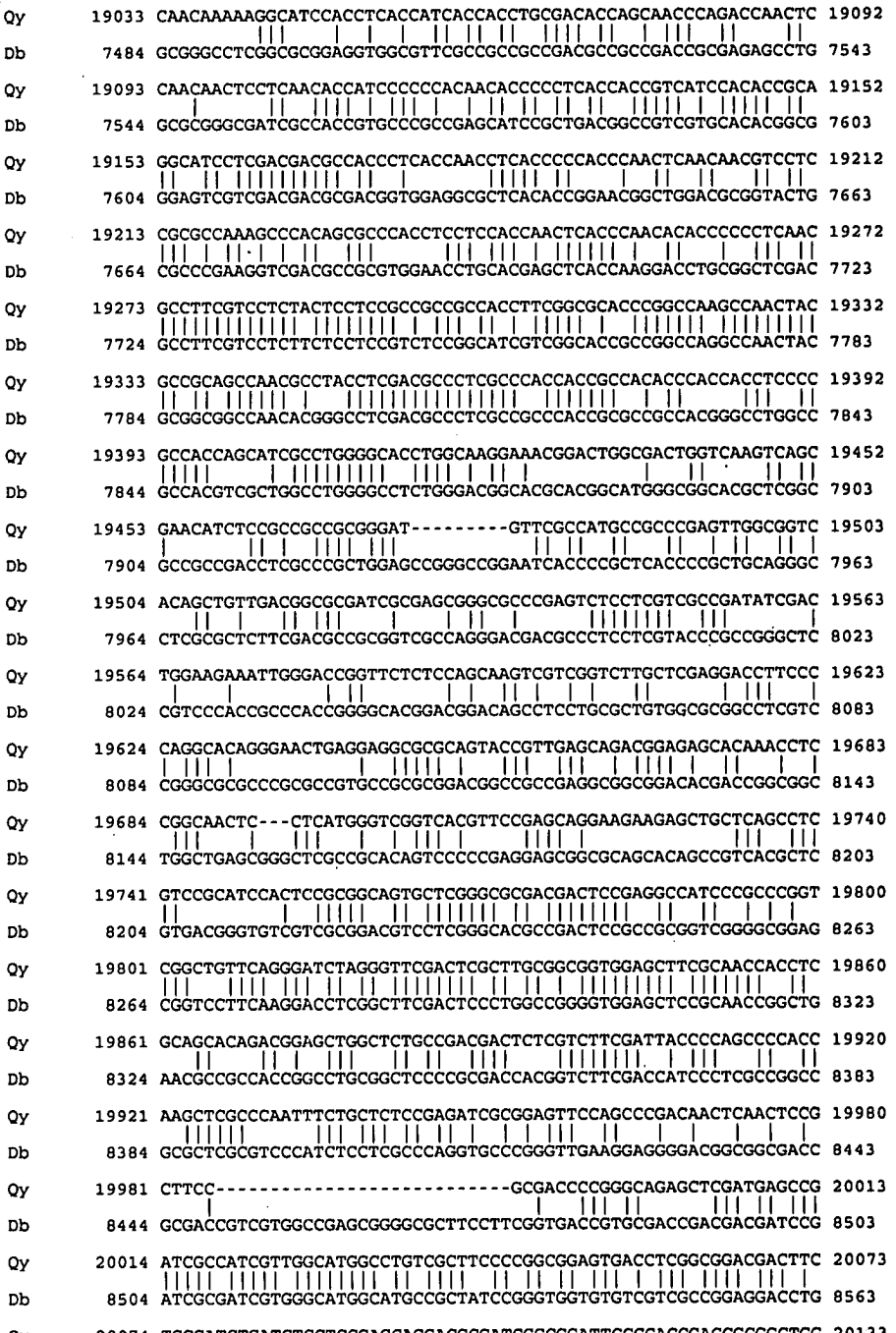
Db       1526 CCGGACCGCGCCCGCCTCCTCGACGGCCTCGGCGCCCTCGCCGCCGGGACGCCCGCGCCC 1585
```

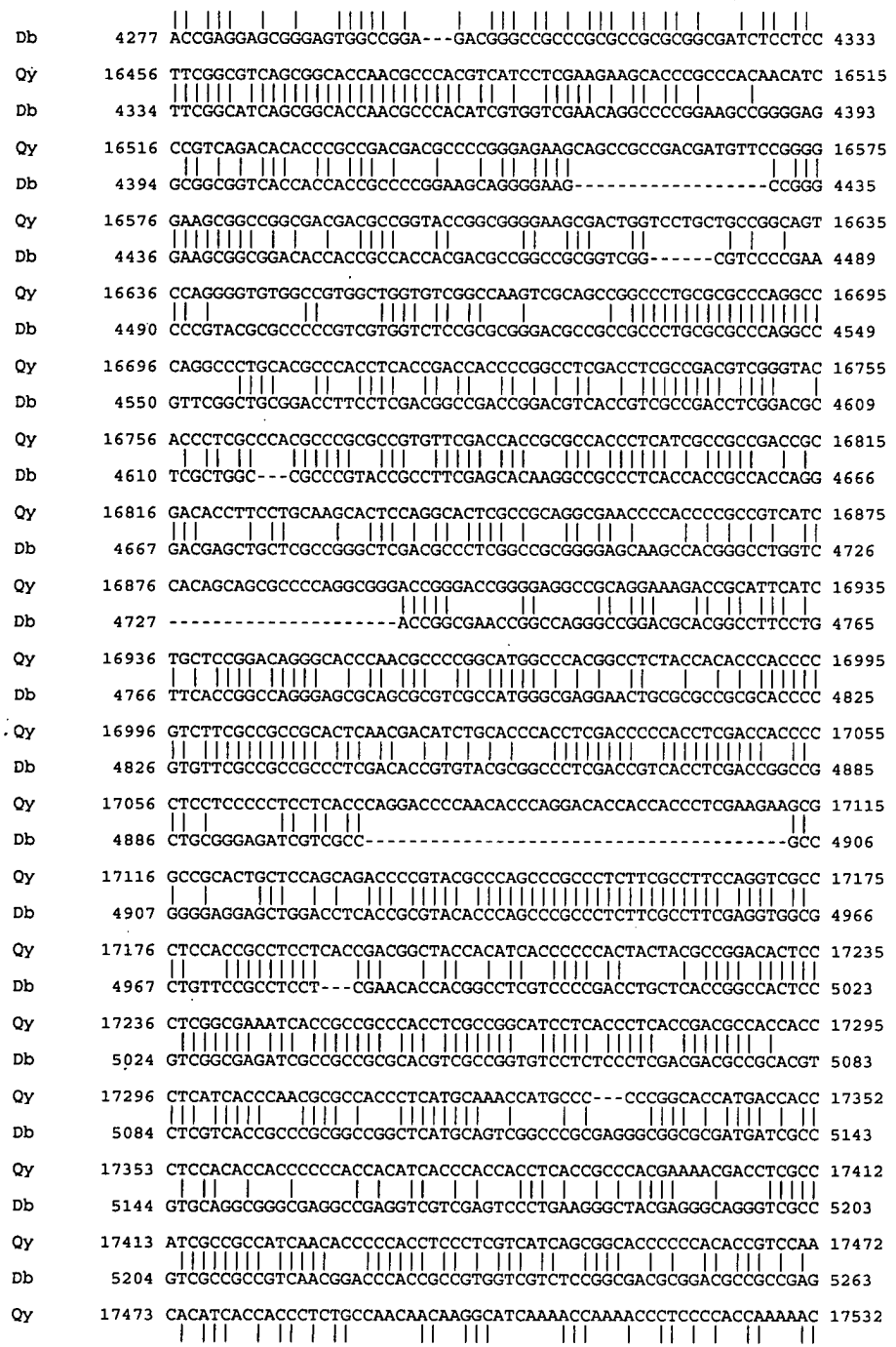

Qy 22291 CAACTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACC 22350
Db 10736 GGATTCCGGGAGGTCGCCGCCGGCCTGCGCTACCGGGAACCGGAGCTGACGGTCGTCTCC 10795
Qy 22351 GCCAACA----------CCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAA 22398
Db 10796 ACGGTCACGGGGCGGCCCGCCCGCCCCGGTGAACTCACCGGCCCCGACTACTGGGTGGCC 10855
Qy 22399 CAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTC 22458
Db 10856 CAGGTCCGTGAGCCCGTGCGCTTCGCGGACGCGGTCCGCACGGCACACCGCCTCGGAGCC 10915
Qy 22459 ACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTC 22518
Db 10916 CGCACCTTCCTGGAGACCGGCCCGGACGGCGTGCTGTGCGGCATGGCAGAGGAGTGCCTG 10975
Qy 22519 CCCAACACCCCCACCACCACCCTCACCCTCACCCACCCCCACCACCACCCCCAAACCCAC 22578
Db 10976 GAGGACGACACCGTGGCCCTGCTGCCGGCGATCCACAAGCCCGGCACCGCG--------- 11026
Qy 22579 CTCCTCACCAACCTCGCCAAAACCACCACCACCTGGCACCCCCACCACTACACCCACCAC 22638
Db 11027 ---------- 11026
Qy 22639 CACAACCAACCCCACACCCACACCCACCTCGACCTCCCCACCTACCCCTTCCAACACCAC 22698
Db 11027 ---------- 11026
Qy 22699 CACTACTGGCTCGAACTACCCAGCGCCCAAACCAGCCCCGGTCAAAGGCGTTCTCGCCGC 22758
Db 11027 ----------CCGCACGGTCCGGCGGCTCCCGGCGCGCTGCGGGCGGCCGCC 11068
Qy 22759 TCGGCTCCAGACACCGCCGAGTCGGAGTTCTGGGACGCGGTGAACGAGGAAGACCTCCAG 22818
Db 11069 GCCGCGTACGGCCGGGGCGCCCGGGTGGACTGGGCCGGGATGCACGCCGACGGCCCCGAG 11128
Qy 22819 AGCCTCGCCGAAACCCTCGACATCGACGCCTCTGCTCTGGACACGGTGGTGCCCGCACTC 22878
Db 11129 GGGCCGGCCCGCCGCGTCGAACTGCCCGTCCACGCCTT----------CCGGCAC 11173
Qy 22879 TCCGCCTGGCACCGCCACCAACACGACCAAGCCCGCATCAACACCTGGACCTACCAGGAA 22938
Db 11174 CGCCGCTACTGGCTCGCCCCGGGCCGCGCGGCGGACACCGACGACTGGATGTACCGGATC 11233
Qy 22939 ACCTGGAAACCCCT---CACCCTCCCCACCACCCACCAACCCCACCAAACCTGGCTCATC 22995
Db 11234 GGCTGGGACCGGCTGCCGGCTGTGACCGGCGGGGCCCGGACCGCCGGCCGCTGGCTGGTG 11293
Qy 22996 GCCATCCCCGAAACCCAGACCCACCACCCCCACATCACCAACATCCTCACCAACCTCCAC 23055
Db 11294 ATCCACCCCGACAGCCCGCGCTGCCGGGAGCTGTCCGGCCACGCCGAACGCGCGCTGCGC 11353
Qy 23056 CACCACGGCATCACCCCCATCCCCCTCACTGTCAACCACACCCACACCAACCCCCAACAC 23115
Db 11354 GCCGCGGGCGCGAGCCCCGTACCGCTGCCCGTGGACGCTCCGGCCGCCGACCGGGCGTCC 11413
Qy 23116 CTCCACCACACCCTCCACCA---CACCCGACAACAAGCCCAAAACCACACCACCGGACCC 23172
Db 11414 TTCGCGGCACTGCTGCGCTCCGCCACCGGACCTGACACACGAGGTGACACAGCCGCGCCC 11473
Qy 23173 ATCACCGGCCTGCTCTCCCTCCTCGCCCTCGACGAAACACCCCACCCCCCACCACCCCCAC 23232
Db 11474 GTGGCCGGTGTGCTGTCGCTGCTGTCCGAGGAGGATCGGCCCCATCGCCAGCACGCCCCG 11533
Qy 23233 ACACCCACCGGCACCCTCCTCAACCTCACCCTCCCCCAAACCCACACCCAAACCCACCCA 23292
Db 11534 GTACCCGCCGGGGTCCTGGCGACGCTGTCCCTGATGCAGGCTATGGAGGAGGAGGCGGTG 11593
Qy 23293 CCAACCCCCCTCTGGTACGCCACCACCAACGCCACCACCACCCACCCCAACGACCCCCTC 23352
Db 11594 GAGGCTCGCGTGTGGTGCGTCTCCCGCGCCGCGGTCGCCGCCGCCGACCGGGAACGGCCC 11653

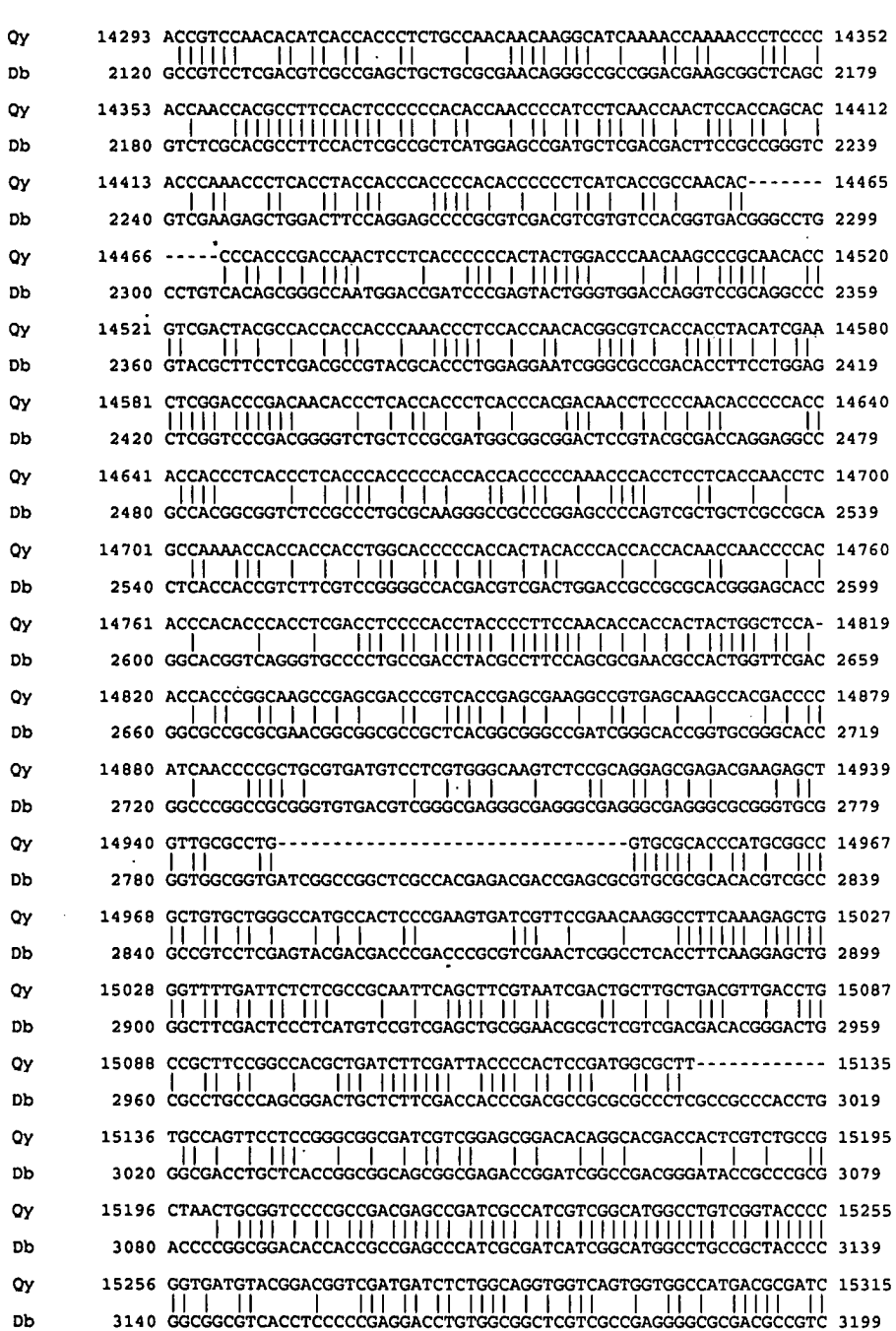

Qy 23353 ACACACCCCACCCAAGCCCAAACCTGGGGACTCGCCCGCACCACCCTCCTCGAACACCCC 23412
Db 11654 GTCGGCGCGGGCGCCGCCC---TGTGGGGGCTGGGGCGGGTGGCCGCCCTGGAACGCCCC 11710
Qy 23413 ACCCACACCGCCGGAATCATCGACCTCCCCACCACCCCCACCCCCCACACCCTCCACCAC 23472
Db 11711 ACCCGGTGGGGCGGTCTCGTGGACCTGCCCGCCTCGCCCGGTGCGGCGCACTGGGCGGCC 11770
Qy 23473 CTCACCCAAACCCTCACCCAACCCCACCACCAAACCCAACTCGCCATCCGCACCACCGGC 23532
Db 11771 GCCGTGGAACGGCTCGCCGGTCCCGAGGACCAGATCGCCGTGCGCGCGTCCGGCAGTTGG 11830
Qy 23533 ACCCACACCCGCCGCCTCACCCCCACCACCCTCACCCCCACACACCAACCACCCACCCCC 23592
Db 11831 GGCCGGCGCCTCACCAGGCTGCCGCGCGACGGCGGCGGCCGGACGGCCGCACCCGCGTAC 11890
Qy 23593 ACCCCCCACGGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCACCTCACC 23652
Db 11891 CGGCCGCGCGGCACGGTGCTCGTCACCGGTGGCACCGGCGCGCTCGGCGGGCATCTCGCC 11950
Qy 23653 CACCACCTCACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACCGGCCCC 23712
Db 11951 CGCTGGCTCGCCGC---GGCGGGCGCCGAACACCTGGCGCTCACCAGCCGCCGGGGCCCG 12007
Qy 23713 CACACCCCCCACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCACCTCACC 23772
Db 12008 GACGCGCCCGGCGCCGCCGGACTCGAGGCCGAACTCCTCCTCCTGGGCGCCAAGGTGACG 12067
Qy 23773 ATCACCACCTGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACCATCCCC 23832
Db 12068 TTCGCCGCCTGCGACACCGCCGACCGCGACGGCCTCGCCCGGGTCCTGCGGGCGATACCG 12127
Qy 23833 CCACAACACCCCCTCACCACCGTCATCCACACCGCAGGCGTCAATCTCTTCGCCCCCGTG 23892
Db 12128 GAGGACACCCCGCTCACCGCGGTGTTCCACGCCGCGGGCGTACCGCAGGTCACGCCGCTG 12187
Qy 23893 TCGGAAACCGATGCCGAATCCTTCTCTTCCGTTACGGCAGCGAAGGCAACGGGCGCGGCG 23952
Db 12188 TCCCGTACCTCGCCCGAGCACTTCGCCGACGTGTACGCGGGCAAGGCGGCGGGCGCCGCG 12247
Qy 23953 ATTCTGCATGAGTTGCTGCTGGACCATGAAAC---GCTTGAACACTTCATTCTCTTCTCG 24009
Db 12248 CACCTGGACGAACTGACCCGCGAACTCGGCGCCGGACTCGACGCGTTCGTCCTCTACTCC 12307
Qy 24010 TCGGGCGCCGGCGCTTGGGGCAGCGGGAATCAGTGCGCATACTCGGCGGCCAACGCATAC 24069
Db 12308 TCCGGCGCCGGCGTCTGGGGCAGCGCCGGCCAGGGTGCCTACGCCGCCGCCAACGCCGCC 12367
Qy 24070 CTGGACGCGCTCGCGACGCATCGTCAGACACATGGACTTCCCGGGGCATCGATCGCCTGG 24129
Db 12368 CTGGACGCGCTCGCCCGGCGCCGTGCGGCGGACGGACTCCCCGCCACCTCCATCGCCTGG 12427
Qy 24130 GGCCCCTGGGCCGGAAAGGGCATGTCGGCCGGTGATGCGGCTCATGGTTACCTGGAAAAG 24189
Db 12428 GGCGTGTGGGGCGGCGGCGGTATGGGGGCCGACGAGGCGGGCGCGGAGTATCTGGGCCGG 12487
Qy 24190 CGCGGCATTCTGCCGATGGAGCCACGCATGGCGCTCGCGGCATTCCATCGTGCGCGGGCG 24249
Db 12488 CGCGGTATGCGCCCCATGGCACCGGTCTCCGCGCTCCGGGCGATGGCCACCGCCATCGCC 12547
Qy 24250 CAGCGGCCGAATTCCAACCTGATCATCGCGGACATCGACTGGGAGCGCTTCGTCCCCGCC 24309
Db 12548 TCCGG---GGAACCCTGCCCCACCGTCACCCACACCGACTGGGAGCGCTTCGGCGAGGGC 12604
Qy 24310 TTCACCGCTCGACGCCACAGCCCGCTCATCGAGGACATTCCGGAGGTTCGGCAAGCGGCT 24369
Db 12605 TTCACCGCCTTCCGGCCCAGCCCTCTGATCGCGG----------GGCTCGGCACGCCGGG 12654
Qy 24370 CAGGAGCTGGAAGCAGCTGCGTCGACGGCAAAGACGACCACAGCTCAGCCGATTGCGACG 24429
Db 12655 CGGCGGCCG----------GGCGGCGGAGACCCCCGAGGAGGGGAACGCCACC 12697
Qy 24430 TCTCTCCGTGAGCGATTGGCCCGACTGACGTCCTCAAAGCAGAACCAGGTGCTGCTCGGC 24489

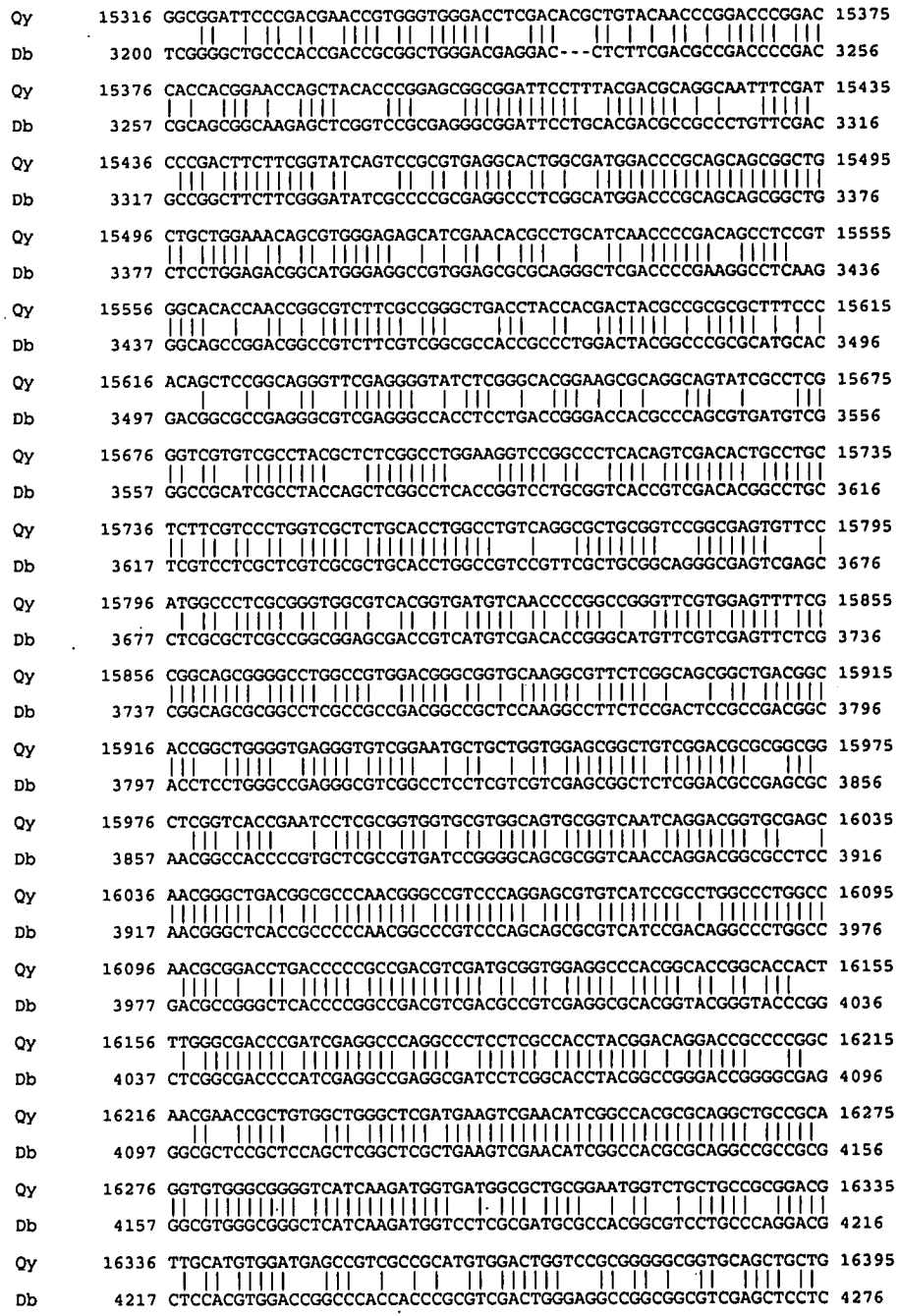

Db 8564 TGGCGGCTGGTGGCCGAGGGGACGGACGCGATCAGCGAGTTCCCCGTCAACCGCGGCTGG 8623
Qy 20134 GACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACCCGAAAC 20193
Db 8624 GACCTGGAGAGCCTCTACGACCCGGATCCCGAGTCGAAGGGCACCACGTACTGCCGGGAG 8683
Qy 20194 GGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGCCCCCGC 20253
Db 8684 GGCGGGTTCCTGGAAGGCGCCGGTGACTTCGACGCCGCCTTCTTCGGCATCTCGCCGCGC 8743
Qy 20254 GAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAAACCATC 20313
Db 8744 GAGGCCCTGGTGATGGACCCGCAGCAGCGGCTGCTGCTGGAGGTGTCCTGGGAGGCGCTG 8803
Qy 20314 GAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTCACCGGC 20373
Db 8804 GAACGCGCGGGCATCGACCCGTCCTCGCTGCGCGGCAGCCGCGGTGGTGTCTACGTGGGC 8863
Qy 20374 ACCAACGGACAGGACCACGCGGCACACATCCGTCAGGCCCCGAGCGGTACCGAGGGATTC 20433
Db 8864 GCCGCGCACGGCTCGTACGCCTCCGATCCCCGGCTGGTGCCCGAGGGCTCGGAGGGCTAT 8923
Qy 20434 GTCCTGACCGGGGCAGCCACCAGCATCGCCTCCGGCCGAATCTCCTACATCCTCGGGTTG 20493
Db 8924 CTGCTGACCGGCAGCGCCGACGCGGTGATGTCCGGCCGCATCTCCTACGCGCTCGGTCTC 8983
Qy 20494 GAAGGGCCTGCGGTCACCCTCGACACAGCGTGTTCCTCCTCGCTCGTCGCCCTGCACCTC 20553
Db 8984 GAAGGACCGTCCATGACGGTGGAGACGGCCTGCTCCTCCTCGCTGGTGGCGCTGCATCTG 9043
Qy 20554 GCCTGCCAGTCCCTCAGGTCCGGTGAATGCACCATGGCCTTGGCCGGCGGGGCCACGGTC 20613
Db 9044 GCGGTACGGGCGCTGCGGCACGGCGAGTGCGGGCTCGCGCTGGCGGGCGGGGTGGCGGTG 9103
Qy 20614 ATGACCACCCCGATCACCTTCACCGAATTCGCCCGCCAACGCGGACTCGCCCCCGACGGG 20673
Db 9104 ATGGCCGATCCCGGCGGCGTTCGTGGAGTTCTCCCGGCAGAAGGGGCTGGCCGCCGACGGC 9163
Qy 20674 CGTTGCAAGGCGTTCTCGGCGGCGGCTGACGGTACCGGCTGGGGTGAGGGTGTGGGGATG 20733
Db 9164 CGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGGCTGGGCCGAGGGCGTCGGCGTG 9223
Qy 20734 CTGCTGGTGGAGCGGCTCTCCGACGCCCGCCGCAACGGTCACCGTGTCCTGGCCGTGGTG 20793
Db 9224 CTCGTCCTGGAGCGGCTGTCGGACGCGCGCCGCGCGGGGCACACGGTCCTCGGCCTGGTC 9283
Qy 20794 CGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAACGGGCCC 20853
Db 9284 ACCGGCACCGCGGTCAACCAGGACGGTGCCTCCAACGGGCTGACCGCGCCCAACGGCCCA 9343
Qy 20854 TCCCAGCAGCGCGTCATCCGCCAGGCCCTCGCCAACGCGGACCTGACCCCCGCCGACGTC 20913
Db 9344 GCCCAGCAACGCGTCATCGCCGAGGCGCTCGCCGACGCCGGGCTGTCCCCGGAGGACGTG 9403
Qy 20914 GATGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCCCAGGCC 20973
Db 9404 GACGCGGTCGAGGCGCACGGCACCGGCACCCGGCTCGGCGACCCCATCGAGGCCGGGGCG 9463
Qy 20974 ATCCTCGCGACCTACGGACAGGACCGTCCCGGCAACGGGCCGTTGTGGCTGGGCTCCGTC 21033
Db 9464 CTGCTCGCCGCCTCCGGACGGAACCGTTCCGGCGACCACCCGCTGTGGCTCGGCTCGCTG 9523
Qy 21034 AAGTCCAACGTCGGACACACACAGGCCGCGGCGGGCGTGGCCGGAGTGATCAAGATGGTG 21093
Db 9524 AAGTCCAACATCGGGCATGCCCAGGCCGCCGCCGGTGTCGGCGGCGTCATCAAGATGCTC 9583
Qy 21094 ATGGCCCTCCGCCACCGGACACTCCCACCGACTCTCCACGCGGATGAGCCGTCGCCGCAT 21153
Db 9584 CAGGCGCTGCGGCACGGCTTGCTGCCCCGCACCCTCCACGCCGACGAGCCGACCCCGCAT 9643
Qy 21154 GTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCCGGCGGG 21213
Db 9644 GCCGACTGGAGCTCCGGCCGGGTACGGCTGCTCACCTCCGAGGTGCCGTGGC---AGCGG 9700
Qy 21214 GAGGGGCGGCCGCGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAACGCCCAC 21273
Db 9701 ACCGGCCGGCCCCGGCGGACCGGGGTGTCCGCCTTCGGCGTCGGCGGCACCAATGCCCAT 9760
Qy 21274 GTCATCCTCGAAGAAGCACCCGCCGACGACGTTCCGGGGGGACCACCCGCCGACGAGGAT 21333
Db 9761 GTCGTCCTCGAAGAGGCACCCGCC---------CCGCCCGCGCCGGAACCGGCCGGGGAG 9811
Qy 21334 GCCGGTAGTGGCGAGGAGGCTGCTGCCGGCAGTCCTGGGGTGTGGCCGTGGCTGGTGTCG 21393
Db 9812 GCCCCCGGCGGCTCCCGCGCCGCAGAAGGGGCGGAAGGGCCCCTGGCCTGGGTGGTCTCC 9871
Qy 21394 GCCAAGTCGCAGCCGGCCCTGCGCGCCCAGGCCCAGGCCCTGCACGCCCACCTCACCGAC 21453
Db 9872 GGACGCGACGAGCCGGCCCTGCGGTCCCAGGCCCGGCGGCTCCGCGACCACCTCTCCCGC 9931
Qy 21454 CACCCCGGCCTCGACCTCGCCGACGTCGGATACACCCTCGCCCACGCCCGCGCCGTGTTC 21513
Db 9932 ACCCCCGGGGCCCGCCCGCGTGACATCGCCTTCTCCCTCGCCGCCACGCGCGCAGCCTTT 9991
Qy 21514 GACCACCGCGCCCACCCTCATCGCCGCCGACCGCGACACCTTCCTGCAAGCACTCCAGGCA 21573
Db 9992 GACCACCGCGCCGTGCTGATCGGCTCGGACGGGGCCGAACTCGCCGCCGCCCTGGACGCG 10051
Qy 21574 CTCGCCGCAGGCGAACCCCACCCCGCCGTCATCCACAGCAGCGCCCCAGGCGGGACCGGG 21633
Db 10052 TTGGCCGAAGG--------ACGCGACGGTCCGGCGGTGGTGCGCGGAGTCCGCGACCGGG 10103
Qy 21634 ACCGGGGAGGCCGCAGGAAAGACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCC 21693
Db 10104 AC-------------GGCAGGATGGCCTTCCTCTTCACCGGGCAGGGCAGCCAGCGCGCC 10150
Qy 21694 GGCATGGCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATC 21753
Db 10151 GGGATGGCCCACGACCTGCATGCCGCCCATACCTTCTTCGCGTCCGCCCTCGACGAGGTG 10210
Qy 21754 TGCACCCACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGAC 21813
Db 10211 ACGGACCGTCTCGACCCGCTGCTCGGCCGGCCGCTCGGCGCGCTGCT----------G 10258
Qy 21814 AACGACAACGACAACGAGGACGCGGCCGCACTGCTCCAGCAGACCCCGTACGCCCAGCCC 21873
Db 10259 GACGCCCGACCCGGCTCGCCCGAAGCGGCACTCCTGGACCGGACCGAGTACACCCAGCCG 10318
Qy 21874 GCCCTCTTCGCCTTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACC 21933
Db 10319 GCGCTCTTCGCCGTCGAGGTGGCGCTCCACCGGCTGCTG---GAGCACTGGGGGATGCGC 10375
Qy 21934 CCCCACTACTACGCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATC 21993
Db 10376 CCCGACCTGCTGCTGGGGCACTCGGTGGGCGAACTGGCGGCCGCCCACGTCGCGGGTGTG 10435
Qy 21994 CTCACCCTCACCGACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATG 22053
Db 10436 CTCGATCTCGACGACGCCTGCGCGCTGGTGGCCGCCCGCGGCAGGCTGATGCAGCGCCTG 10495
Qy 22054 CCCCC---CGGCACCATGACCACCCTCCACACCACCCCACACCACATCACCCACCACCTC 22110
Db 10496 CCGCCCGGCGGCGCGATGGTCTCCGTGCGGGCCGGCGAGGACGAGGTCCGCGCACTGCTG 10555
Qy 22111 ACCGCCCACGAAAACGACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATC 22170
Db 10556 GCCGGCCGCGAGGACGCCGTCTGCGTCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATC 10615
Qy 22171 AGCGGCACCCCCCACACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAA 22230
Db 10616 TCCGGCGCGGAGGAAGCGGTGGCCGAGGCGGCGGCGCAGCTCGCCGGACGAGGCCGCCGC 10675
Qy 22231 ACCAAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAAC 22290
Db 10676 ACCAGGCGGCTCCGCGTCGCGCACGCCTTCCACTCACCCCTGATGGACGGCATGCTCGCC 10735






Db 4277 ACCGAGGAGCGGGAGTGGCCGGA---GACGGGCCGCCCGCGCCGCGCGGCGATCTCCTCC 4333
Qy 16456 TTCGGCGTCAGCGGCACCAACGCCCACGTCATCCTCGAAGAAGCACCCGCCCACAACATC 16515
Db 4334 TTCGGCATCAGCGGCACCAACGCCCACATCGTGGTCGAACAGGCCCCGGAAGCCGGGGAG 4393
Qy 16516 CCGTCAGACACACCCGCCGACGACGCCCCGGGAGAAGCAGCCGCCGACGATGTTCCGGGG 16575
Db 4394 GCGGCGGTCACCACCACCGCCCCGGAAGCAGGGGAAG----------CCGGG 4435
Qy 16576 GAAGCGGCCGGCGACGACGCCGGTACCGGCGGGGAAGCGACTGGTCCTGCTGCCGGCAGT 16635
Db 4436 GAAGCGGCGGACACCACCGCCACCACGACGCCGGCCGCGGTCGG------CGTCCCCGAA 4489
Qy 16636 CCAGGGGTGTGGCCGTGGCTGGTGTCGGCCAAGTCGCAGCCGGCCCTGCGCGCCCAGGCC 16695
Db 4490 CCCGTACGCGCCCCCGTCGTGGTCTCCGCGCGGGACGCCGCCGCCCTGCGCGCCCAGGCC 4549
Qy 16696 CAGGCCCTGCACGCCCACCTCACCGACCACCCCGGCCTCGACCTCGCCGACGTCGGGTAC 16755
Db 4550 GTTCGGCTGCGGACCTTCCTCGACGGCCGACCGGACGTCACCGTCGCCGACCTCGGACGC 4609
Qy 16756 ACCCTCGCCCACGCCCGCGCCGTGTTCGACCACCGCGCCACCCTCATCGCCGCCGACCGC 16815
Db 4610 TCGCTGGC---CGCCCGTACCGCCTTCGAGCACAAGGCCGCCCTCACCACCGCCACCAGG 4666
Qy 16816 GACACCTTCCTGCAAGCACTCCAGGCACTCGCCGCAGGCGAACCCCACCCCGCCGTCATC 16875
Db 4667 GACGAGCTGCTCGCCGGGCTCGACGCCCTCGGCCGCGGGGAGCAAGCCACGGGCCTGGTC 4726
Qy 16876 CACAGCAGCGCCCCAGGCGGGACCGGGACCGGGGAGGCCGCAGGAAAGACCGCATTCATC 16935
Db 4727 ----------ACCGGCGAACCGGCCAGGGCCGGACGCACGGCCTTCCTG 4765
Qy 16936 TGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCACACCCACCCC 16995
Db 4766 TTCACCGGCCAGGGAGCGCAGCGCGTCGCCATGGGCGAGGAACTGCGCGCCGCGCACCCC 4825
Qy 16996 GTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTCGACCACCCC 17055
Db 4826 GTGTTCGCCGCCGCCCTCGACACCGTGTACGCGGCCCTCGACCGTCACCTCGACCGGCCG 4885
Qy 17056 CTCCTCCCCCTCCTCACCCAGGACCCCAACACCCAGGACACCACCACCCTCGAAGAAGCG 17115
Db 4886 CTGCGGGAGATCGTCGCC----------GCC 4906
Qy 17116 GCCGCACTGCTCCAGCAGACCCCGTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCC 17175
Db 4907 GGGGAGGAGCTGGACCTCACCGCGTACACCCAGCCCGCCCTCTTCGCCTTCGAGGTGGCG 4966
Qy 17176 CTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCC 17235
Db 4967 CTGTTCCGCCTCCT---CGAACACCACGGCCTCGTCCCCGACCTGCTCACCGGCCACTCC 5023
Qy 17236 CTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACC 17295
Db 5024 GTCGGCGAGATCGCCGCCGCGCACGTCGCCGGTGTCCTCTCCCTCGACGACGCCGCACGT 5083
Qy 17296 CTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCC---CCCGGCACCATGACCACC 17352
Db 5084 CTCGTCACCGCCCGCGGCCGGCTCATGCAGTCGGCCCGCGAGGGCGGCGCGATGATCGCC 5143
Qy 17353 CTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCC 17412
Db 5144 GTGCAGGCGGGCGAGGCCGAGGTCGTCGAGTCCCTGAAGGGCTACGAGGGCAGGGTCGCC 5203
Qy 17413 ATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAA 17472
Db 5204 GTCGCCGCCGTCAACGGACCCACCGCCGTGGTCGTCTCCGGCGACGCGGACGCCGCCGAG 5263
Qy 17473 CACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAAAAAC 17532
Db 5264 GAGATCCGCGCCGTATGGGCGGGACGCGGCCGGCGCACCCGCAGGCTGCGCGTCAGCCAC 5323
Qy 17533 GCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACC 17592
Db 5324 GCCTTCCACTCCCCGCACATGGACGACGTCCTCGACGAGTTCCTCCGGGTCGCCGAGGGC 5383
Qy 17593 CTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACAC----------CCCA 17640
Db 5384 CTGACCTTCGAGGAGCCGCGGATCCCCGTCGTCTCCACGGTCACCGGCGCGCTCGTCACG 5443
Qy 17641 CCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTAC 17700
Db 5444 TCCGGCGAGCTCACCTCGCCCGCGTACTGGGTCGACCAGATCCGGCGGCCCGTGCGCTTC 5503
Qy 17701 GCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCC 17760
Db 5504 CTGGACGCCGTCCGCACCCTGGCCGCCCAGGACGCGACCGTCCTCGTCGAGATCGGCCCC 5563
Qy 17761 GACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACACCCCCACCACCAC---- 17816
Db 5564 GACGCCGTCCTCACGGCACTCGCCGAGGAGGCTCTCGCGCCCGGCACGGACGCCCCGGAC 5623
Qy 17817 --------CCTCACCCTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTCACCAAC 17868
Db 5624 GCCCGGGACGTCACGGTCGTCCCGCTGCTGCGCGCGGGGCGCCCCGAGCCCGAGACCCTC 5683
Qy 17869 CTCGCCAAAACCACCACCACCTGGCACCCCCACCACTACACCCACCACCACAACCAACCC 17928
Db 5684 GCCGCCGGTCTCGCGACCGCCCATGTCCACGGCGCACCCTTGGACCGGGCGTCGTTCTTC 5743
Qy 17929 CACACCCACACCCACCTCGACCTCCCCACCTACCCCTTCCAACACCAGCACTACTGGCTC 17988
Db 5744 CCGGACGGGCGCCGCACGGACCTGCCCACGTACGCCTTCCGGCGCGAGCACTACTGGCTG 5803
Qy 17989 GAAAGCACACAGCCGGGTGCCGGATCCGGTTCGGGTTCCGGTTCCGGGCGGGCAGGGACT 18048
Db 5804 ACGCCCGAGGCCCGTACGGACGCCCGCGCACTCGGCTTCGACCCGGCGCGGCACCCGCTG 5863
Qy 18049 GCGGGCGGGACGGCAGAGGTGGAGTCGCGGTTCTGGGACGCGGTGGCCCGCCAGGACCTG 18108
Db 5864 CTGACGACCACGGTCGAGGTCGCCGGCGGCGACGGCGTCCTGCTGACCGGCCGTCTCTCC 5923
Qy 18109 GAAACGGTCGCGACCACGCTCGCCGTGCCCCCCTCCGCCGGCCTGGACACGGTGGTGCCC 18168
Db 5924 CTGACCGACCAGCCCTGGCTGGCCGACCACATGGTCAACGGCGCCGTCCTGTTGCCGGCC 5983
Qy 18169 GCACTCTCC---------- 18177
Db 5984 ACCGCCTTCCTGGAGCTCGCCCTCGCGGCGGGCGACCACGTCGGGGCGGTCCGGGTGGAG 6043
Qy 18178 ----------GCCTGGCACCGCCACCAACACGACCAAGCC 18207
Db 6044 GAACTCACCCTCGAAGCGCCGCTCGTCCTGCCCGAGCGGGGCGCCGTCCGCATCCAGGTC 6103
Qy 18208 CGCATCAACACCTGGACCTACCAGGAAACCTGGAAACCCCTCACCCTCCCCACCACCCAC 18267
Db 6104 GGCGTGAGCGGCGACGGCGAGTCGCCGGCCGGGCGCACCTTCGGTGTGTACAGCACCCCC 6163
Qy 18268 CAACCCCACCAAACCTGGCTCATCGCCATCCCCGAA---------- 18303
Db 6164 GACTCCGGCGACACCGGTGACGACGCGCCCCGGGAGTGGACCCGCCATGTCTCCGGCGTA 6223
Qy 18304 ---------- 18303
Db 6224 CTCGGCGAAGGGGACCCGGCCACGGAGTCGGACCACCCCGGCACCGACGGGGACGGTTCA 6283
Qy 18304 ----------ACCCAGACCCACCACCCCCACATCACCAACATCCTC 18339
Db 6284 GCGGCCTGGCCGCCTGCGGCGGCGACCGCCACACCCCTCGACGGCGTCTACGACCGGCTC 6343
Qy 18340 ACCAACCTCCACCA---------- 18353
Db 6344 GCGGAGCTCGGCTACGGATACGGTCCGGCCTTCCAGGGCCTGACGGGGCTGTGGCGCGAC 6403





```
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 1
;   LENGTH: 15872
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-1

  Query Match             7.1%;  Score 2188.4;  DB 9;  Length 15872;
  Best Local Similarity   52.5%;  Pred. No. 0;
  Matches 6973;  Conservative    0;  Mismatches 5061;  Indels 1248;  Gaps   50;

Qy      12256 CGCAGCGGAGGGTTTCTCTATGACGCGGGCGACTTCGACGCCGCGTTCTTCGAGTTGTCA 12315
              ||| | || || || ||| |   |  | | |||||||| || |||||||   | ||
Db        197 CGCCGGGGCGGCTTCCTCGACTCCCTCGACCTCTTCGACGCGGCCTTCTTCGGAATCTCG 256

Qy      12316 CCGCGTGAGGCGCTGGCAATGGACCCGCAGCAGCGCCTGCTGCTCGAAACCACTTGGGAA 12375
              || |||||||||    || ||||||||||||||||||| ||| | ||||||   | | |||||
Db        257 CCCCGTGAGGCCGCCGCCATGGACCCGCAGCAGCGACTGGTCCTCGAACTCGCCTGGGAG 316

Qy      12376 ACGTTCGAACAGGGCGGAATCGACCCGAGGTCCATGCGCGGAAGCCGGACCGGGGTTTTC 12435
              || | ||  | | |||||||| |||   |  || |   |||||||| ||||  || |||
Db        317 GCGCTGGAGGACGCCGGAATCGTCCCCGGCACCCTCGCCGGAAGCCGCACCGCCGTCTTC 376

Qy      12436 GTGGGGATCAATCCGGAGGACTACACCACCGGATACACACATCAGCCCTCAAACGCAGTC 12495
              || || | |   | ||| |||||||| | |     |  || ||  |    | ||  ||
Db        377 GTCGGCACCCTGCGGGACGACTACACGAGCCTCCTCTACCAGCACGGCGAGCAGGCCATC 436

Qy      12496 GAGGGCTACCTGCTCACTGGCAGCGCGGCAAGCATTGCGTCAGGCCGTATCTCCTACAAC 12555
                    ||    |  | |||             || |     |   |||  |||| ||| ||
Db        437 ACCCAGCACACCATGGCGGGCGTGAACCGGGGCGTCATCGCCAACCGCGTCTCGTACCAC 496

Qy      12556 TTCGGGCTCGAAGGCCCTGCGATCACTATCGACACCGCGTGTTCCTCCTCGCTCGTCGCC 12615
              ||||  ||  | |||||      ||||  |||||  |||||   || ||||||||||||||
Db        497 CTCGGCCTGCAGGGCCCGAGCCTCACCGTCGACGCCGCGCAGTCGTCCTCGCTCGTCGCC 556

Qy      12616 CTGCATCTGGCCTGCCAAGCGCTCCGGTCCGGTGAATGCACCATGGCGCTCGCAGGCGGC 12675
              ||||  |||||||||  |   | || ||  |||| || | ||| | ||||||||  | ||||
Db        557 GTGCACCTGGCCTGCGAGTCCCTGCGCGCCGGGGAGTCCACGACGGCGCTCGTCGCCGGC 616

Qy      12676 GCCTCCGTCATGGCCACTCCCTTCGTCTTCACCGAGTTCTCTCGCCAGCGGGGCCTGGCC 12735
              |    | |||      |     |    |  |     |     |||    | || ||  ||
Db        617 GTGAACCTCAACATCCTCGCGGAGAGCGCCGTGACGGAGGAGCGCTTCGGTGGACTCTCC 676

Qy      12736 GCAGACGGCCGGTGCAAGGCGTTTTCGGCGGCGGCGGACGGGACCGGCTGGTCCGAGGGT 12795
              | ||||||      | |  | ||    |||  |||  ||||    || | ||  ||||||
Db        677 CCGGACGGCACCGCCTACACCTTCGACGCGCGGGCCAACGGATTCGTCCGGGGCGAGGGC 736

Qy      12796 GTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCGCCGCAACGGTCACCGTGTCCTG 12855
              |  ||  |  |  |  |  ||| |||||||  |||||  |  | ||||  |||||||||
Db        737 GGCGGAGTCGTCGTACTCAAGCCGCTCTCCCGCGCCCTCGCCGACGGCGACCGTGTCCAC 796

Qy      12856 GCCGTCGTCCGCGGCAGCGCCGTCAACCAGGACGGCGCAAGCAACGGCCTGACCGCACCC 12915
              | |||| |||||| |||||||||||||| | |||||  || | |    ||  ||  |||
Db        797 GGCGTCATCCGCGCCAGCGCCGTCAACAACGACGGAGCCACCCCGGGTCTCACCGTGCCC 856

Qy      12916 AACGGTCGTTCACAAGTCAAGGTCATCCGCCAGGCTTTGGCCAACGCACACCTCTCCCCT 12975
              | | |      | || |  |||||  | ||| |||| |      || ||   |||   |||
Db        857 AGCAGGGCCGCCCAGGAGAAGGTGCTGCGCGAGGCGTACCGGAAGGCGGCCCTGGACCCG 916

Qy      12976 GCCGATGTCGATGCGGTGGAGGCCCACGGCACGGGGACCACCCTGGGCGACCCGATCGAG 13035
              |||  ||| |    || ||   |||||||| || ||| || | |||||||| ||||||
Db        917 TCCGCCGTCCAGTACGTCGAACTCCACGGCACCGGAACCCCCGTCGGCGACCCCATCGAG 976

Qy      13036 GCTCAAGCCCTCGTCGAAGCCTACGGTCAGGACCGCCCCAACGGCCGCCCCCTCTGGCTC 13095
              ||    || |||| ||  | |  |||   ||  ||||||    | |   |||||     ||
Db        977 GCCGCCGCGCTCGGCGCCGTCCTCGGCTCGGCGCGCCCCGCGGACGAACCCCTGCTCGTC 1036

Qy      13096 GGAACCCTCAAGTCCAACATCGGGCACTCCATGGCCGCTGCGGGTGTGGGCGGGGTCATC 13155
              ||   |  |||| | ||| ||||||||  |   | ||| || ||  | | |||  |||||
Db       1037 GGCTCGGCCAAGACGAACGTCGGGCACCTCGAAGGCGCCGCCGGCATCGTCGGCCTCATC 1096

Qy      13156 AAGATGGTGATGGCGCTGCGGAATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCG 13215
              |||| | |  | |||||  |     | | | |||  ||   |  |  |   |  |||

Db      1097 AAGACGCTCCTCGCGCTCGGCCGGCGCCGGATCCCGGCGAGCCTCAACTTCCGTACGCCC 1156

Qy      13216 TCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGG 13275
              |||| |  |     | |  | | |  |  |||    |  |  |  |||
Db      1157 CACCCGGACATCCCGCTCGACACCCTCGGGCTCGACGTGCCCGACGGCCTGCGGGAGTGG 1216

Qy      13276 CCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 13335
              ||   |  |||    ||   ||  |  || || ||     || ||||||  |  ||||||||
Db      1217 CCGCACCCGGAC---CGCGAACTCCTCGCCGGCGTCAGCTCGTTCGGCATGGGCGGCACC 1273

Qy      13336 AACGCCCACGTCATCCTCGAGGAAGCACCCGCCCACAACATCCCGTCAGACACACCCGCC 13395
              ||||||||||| |||||    ||||  ||||||||
Db      1274 AACGCCCACGTCGTCCTCAGCGAAGGCCCCGCCCAG---------------------- 1309

Qy      13396 GACGACGTCCCGGGAGAATCAGCCGCCGACGAGGATGCCGGTAGTGGCGATGAGGCTGCT 13455
                                                | || |||| |  |||| |  |  || ||| |  |
Db      1310 ----------------------GGCGGCGAGCAGCCCGGCATCGATGAGGAGACCCCC 1345

Qy      13456 GCCGGCAGTCCAGGGGTGTGGCCGTGGCTGGTGTCGGCCAAGTCGCAGCCGGCCCTGCGC 13515
              | || |||    |  |    ||| |  | ||  | | |              ||||||||||
Db      1346 GTCGACAGCGGGGCCGCACTGCCCTTCGTCGTCACCGGCCGCGGCGGCGAGGCCCTGCGC 1405

Qy      13516 GCCCAGGCCCAGGCCCTGCACGCCCACCTCACCGACCACCCCGGCCTCGACCTCGCCGAC 13575
              |||||||||| |  ||||||||     | ||  |  |||| |  ||||  | |||||
Db      1406 GCCCAGGCCCGGCGCCTGCACGAGGCCGTCGAAGCGGACCCGGAGCTCGCGCCCGCCGCA 1465

Qy      13576 GTCGGGTACACCCTCGCCCACGCCCGCGCCGTGTTCGACCACCGCGCCACCCTCATCGCC 13635
              |||      | || | |  | |||| | || |||   |||||  |   | || |||||
Db      1466 CTCGCCCGGTCGCTGGTCACCACCCGTACGGTCTTCACGCACCGGTCGGTCGTCCTCGCC 1525

Qy      13636 GCCGACCGCGACACCTTCCTGCAAGCACTCCAGGCACTCGCCGCAGGCGAACCCCACCCC 13695
              | |||||||  |  | |||| |  | ||  || |||||||| ||     |||  |||
Db      1526 CCGGACCGCGCCCGCCTCCTCGACGGCCTCGGCGCCCTCGCCGCCGGGACGCCCGCGCCC 1585

Qy      13696 GCCGTCATCCACAGCAGCGCCCCAGGCGGGACCGGGACCGGGGAGGCCGCAGGAAAGACC 13755
              | |||  ||  | |||  | ||                          |||  | ||      |
Db      1586 GGCGTGGTCACCGGCACCCCC----------------------GCCCCCGGGCGCCTC 1621

Qy      13756 GCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCAC 13815
              ||  || | | |||  |||||  |||||||  | |||||||  |    |  |||  |
Db      1622 GCCGTCCTGTTCAGCGGCCAGGGTGCCCAACGTACGGGCATGGGCATGGAGTTGTACGCC 1681

Qy      13816 ACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTC 13875
              |||||||||  |||||  | ||  || ||| | ||  | ||  | || |||||  ||||
Db      1682 GCCCACCCCGCCTTCGCGACGGCCTTCGACGCCGTCGCCGCCGAACTGGACCCCCTCCTC 1741

Qy      13876 GACCACCCCCTCCTCCCCCTCCTCACCCAGGACCCCAACACCCAGGACACCACCACCCTC 13935
              ||||  |||||                                  | ||   |||
Db      1742 GACCGGCCCCT----------------------------------------CGCCGAACTC 1762

Qy      13936 GAAGAAGCGGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTC 13995
              |  |  |||| ||      ||| | |  |||  | || | |||||||| |||||||| |
Db      1763 GTCGCGGCGGGCGACACCCTCGACCGCACCGTCCACACACAGCCCGCGCTCTTCGCCGTG 1822

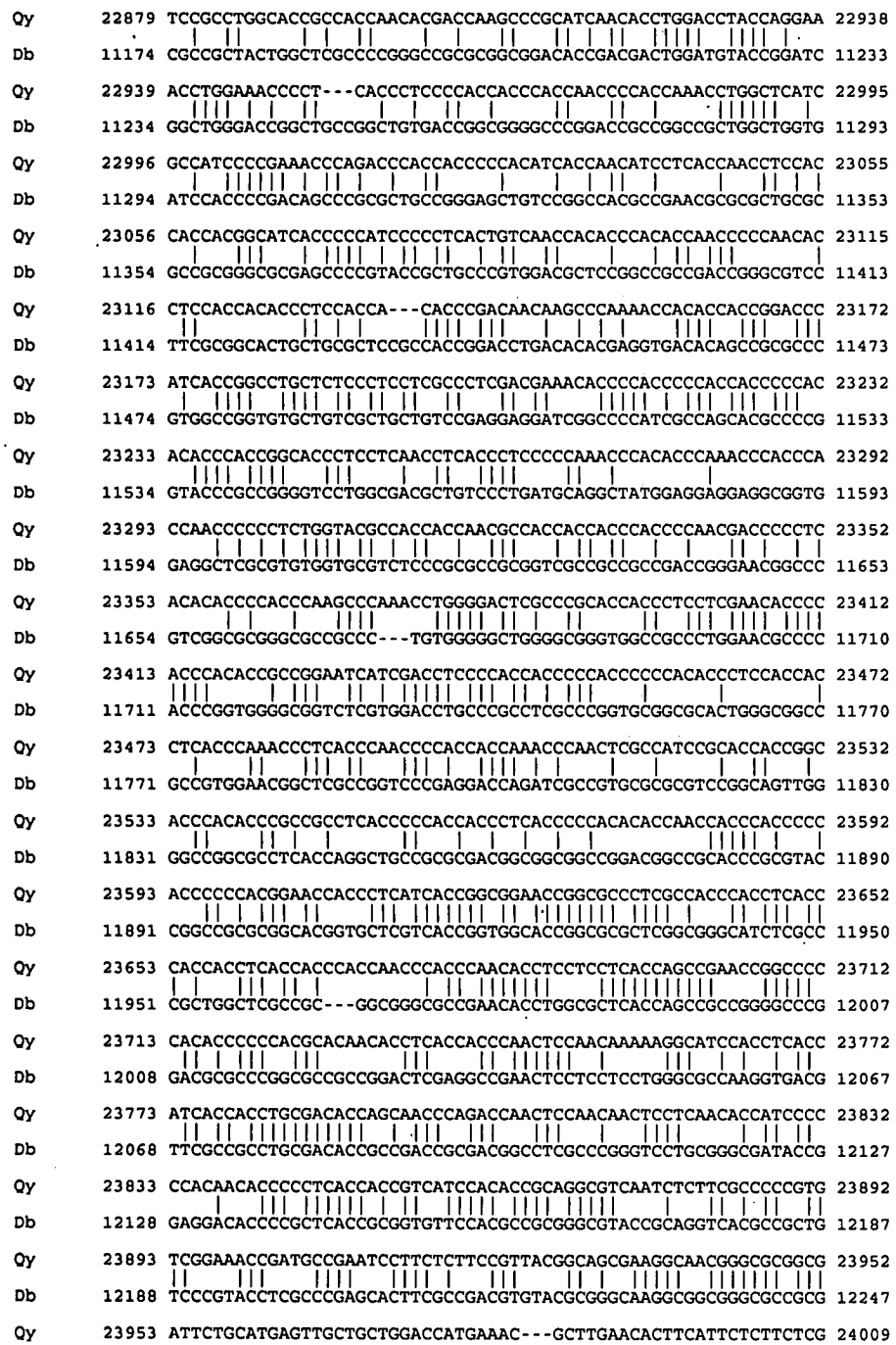
Qy      13996 CAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCC 14055
              |||||||||||||||||||| |  |||  ||    | |||| ||| ||    ||||
Db      1823 GAGGTCGCCCTCCACCGCCTCGT---CGAGTCCTGGGGCGTCACGCCCGACCTGCTCGCC 1879

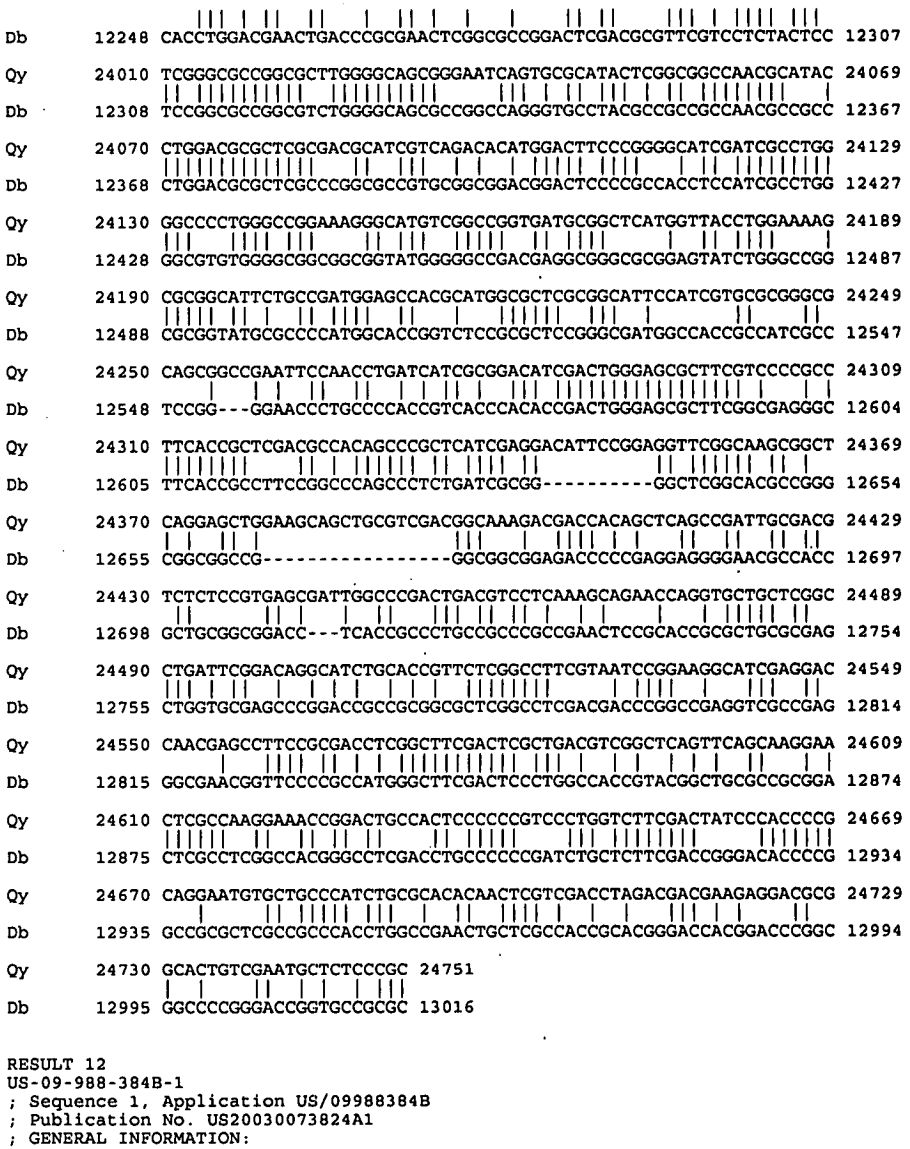
Qy      14056 GGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGAC 14115
              || |||||| ||||||| |||| |||||||||| |||||||  ||||  | ||    ||||
Db      1880 GGCCACTCCGTCGGCGAGATCAGCGCCGCCCACGTCGCCGGGGTCCTGTCGCTGCGCGAC 1939

Qy      14116 GCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCC---CCCGGCACC 14172
              ||| ||  |||| || |    |||| |  |||||||||  |  | |||      |||| |
Db      1940 GCCGCCCGCCTCGTCGCGGCGCGCGGCCGCCTCATGCAGGCGCTCCCCGAGGGCGGCGCG 1999

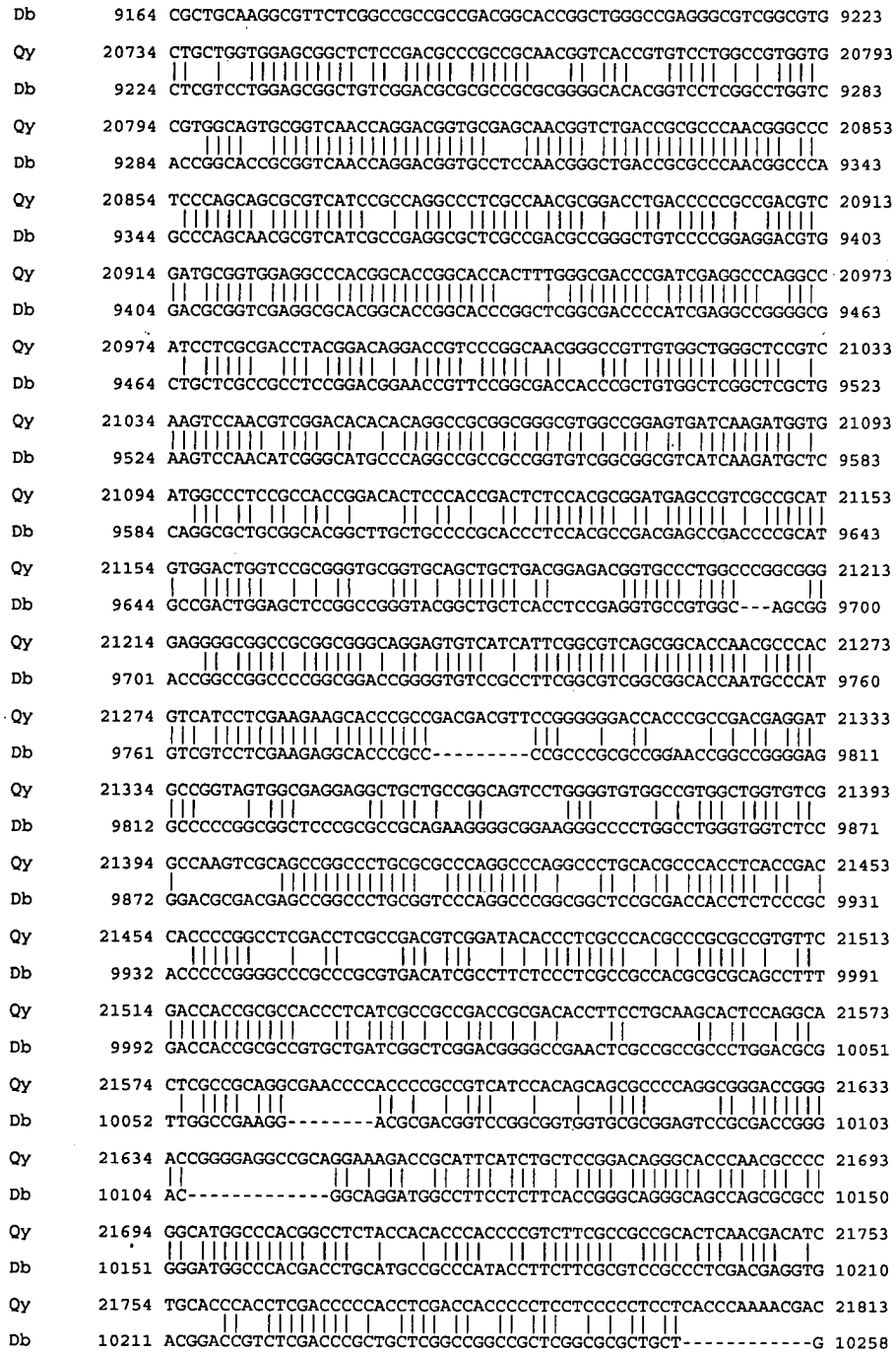
Qy      14173 ATGACCACCCTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAAC 14232
              |||  |  ||  |  |  |  |  |  |      |  ||||||  |  |  |  |||
Db      2000 ATGGTCGCGGTCGAGGCGAGCGAGGAGGAAGTGCTTCCGCACCTCGCGGGACGCGAGCGG 2059

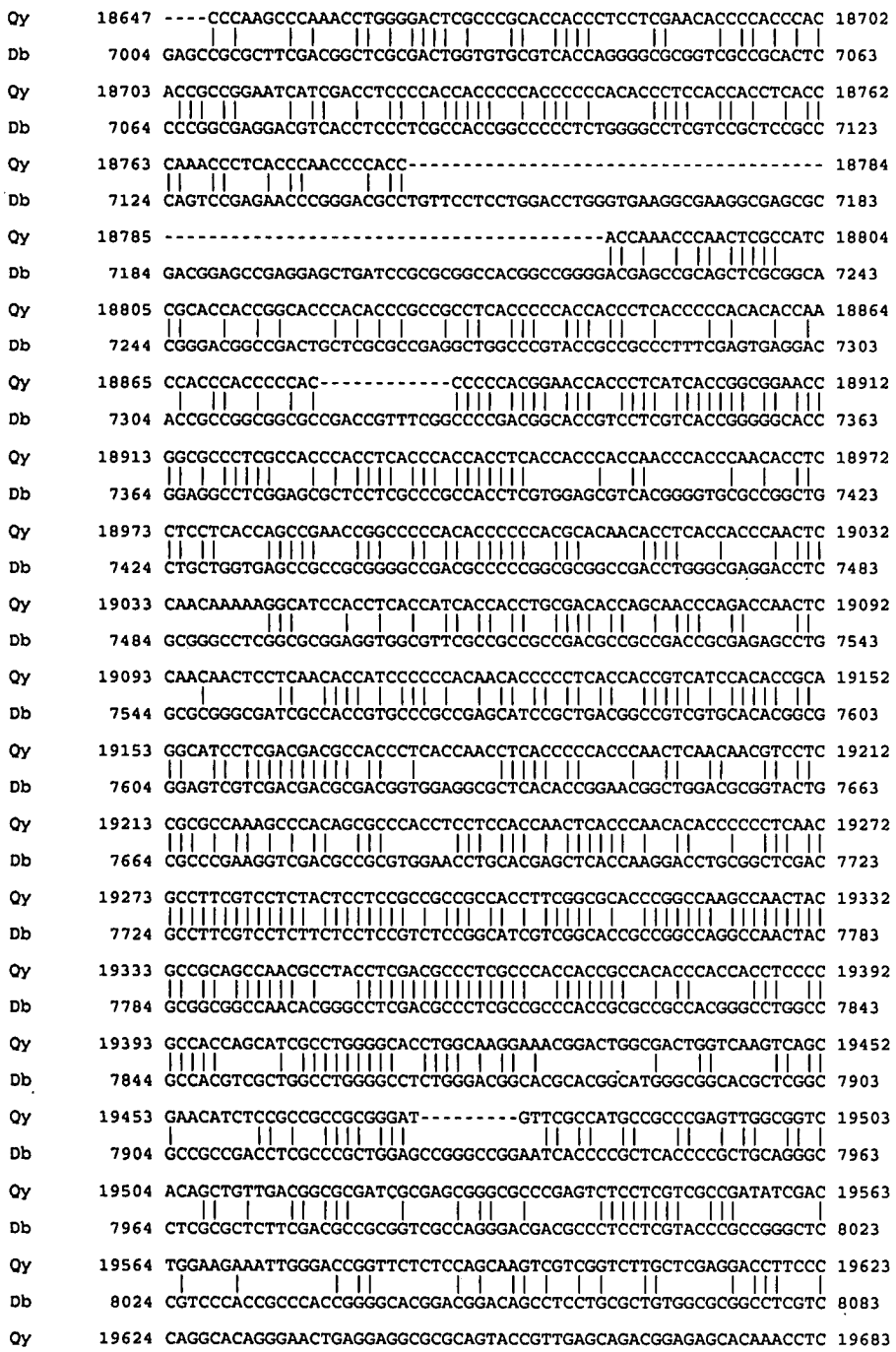
Qy      14233 GACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCAC 14292
              || ||| || |||| ||| | |||  ||||  |  |  |||||  ||   ||| ||   | |
Db      2060 GAGCTCTCCCTCGCGGCCGTGAACGGCCCCCGCGCGGTCGTCCTCGCGGGCGCCGAGCGC 2119
```

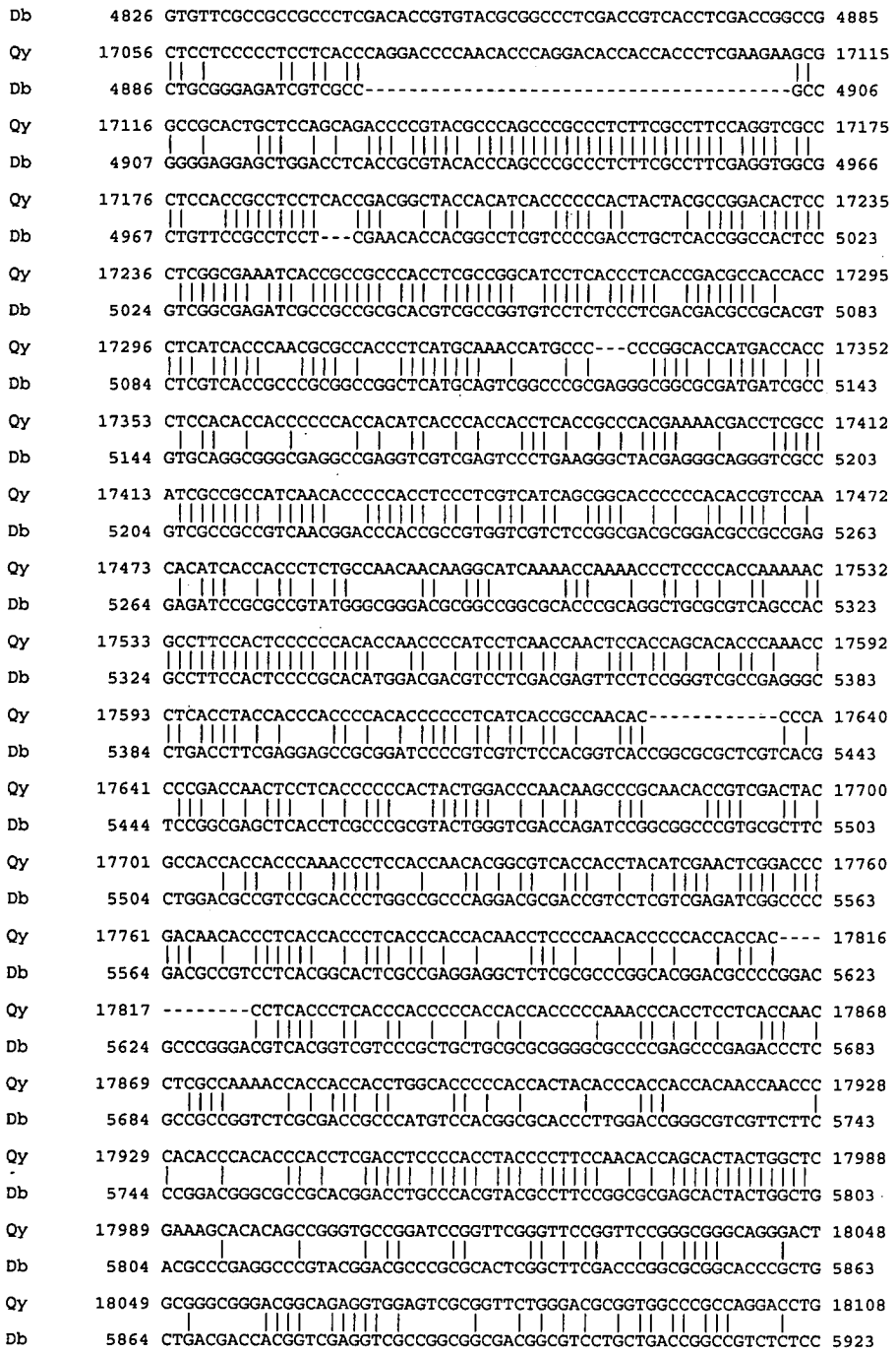


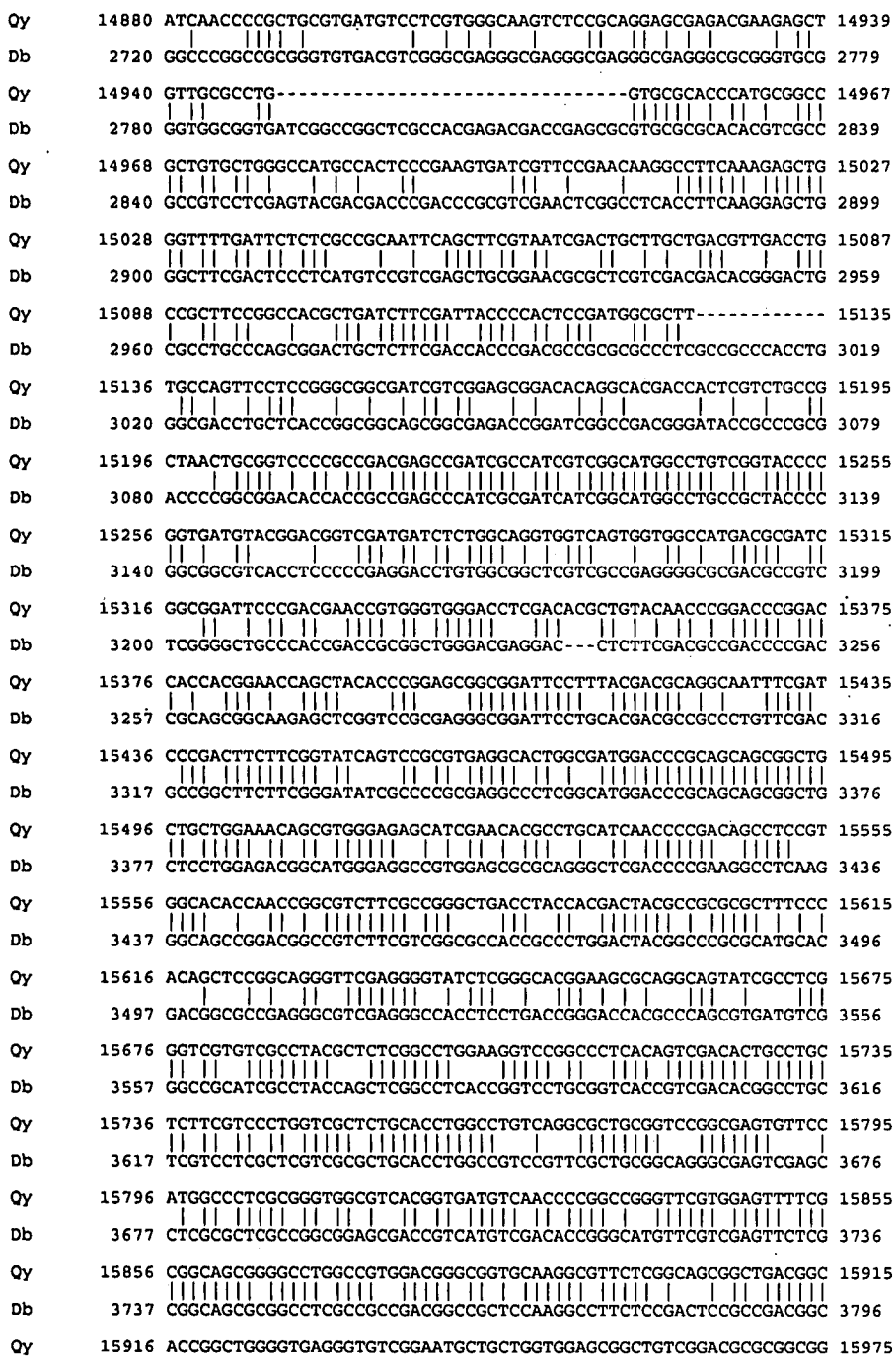


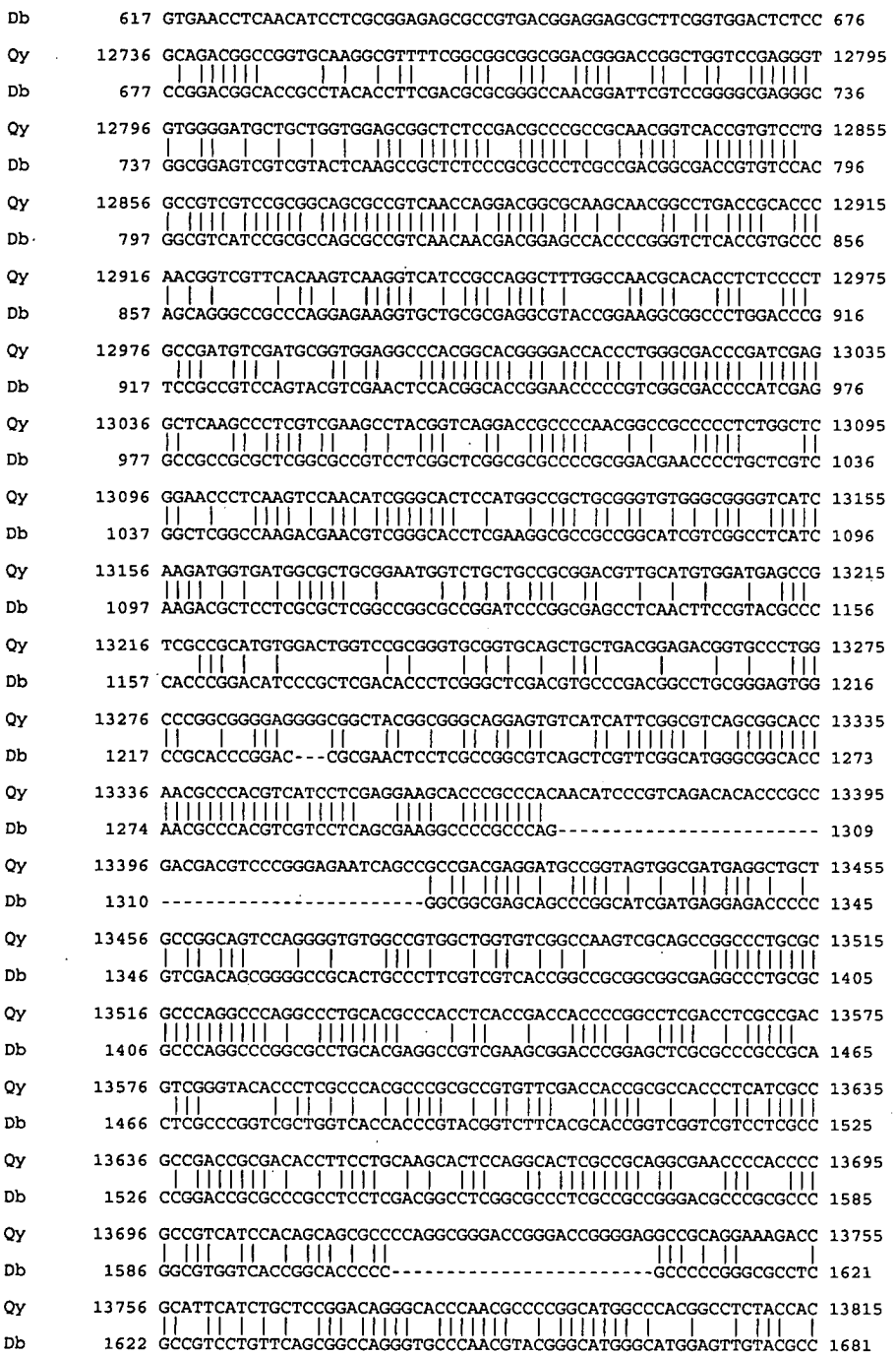
RESULT 12
US-09-988-384B-1
; Sequence 1, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26

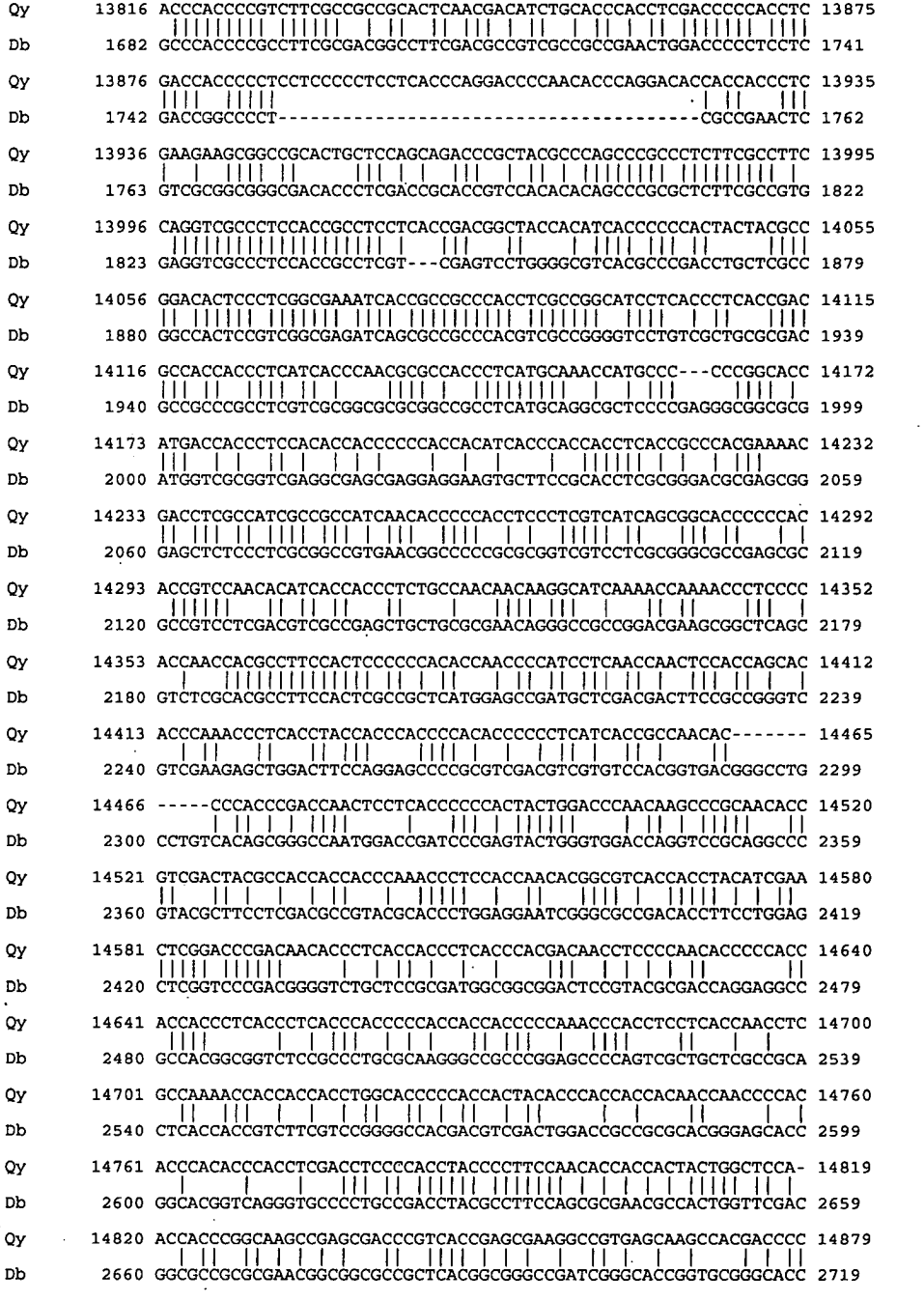

Db 9164 CGCTGCAAGGCGTTCTCGGCCGCCGCCGACGGCACCGGCTGGGCCGAGGGCGTCGGCGTG 9223
Qy 20734 CTGCTGGTGGAGCGGCTCTCCGACGCCCGCCGCAACGGTCACCGTGTCCTGGCCGTGGTG 20793
Db 9224 CTCGTCCTGGAGCGGCTGTCGGACGCGCGCCGCGCGGGGCACACGGTCCTCGGCCTGGTC 9283
Qy 20794 CGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCGCCCAACGGGCCC 20853
Db 9284 ACCGGCACCGCGGTCAACCAGGACGGTGCCTCCAACGGGCTGACCGCGCCCAACGGCCCA 9343
Qy 20854 TCCCAGCAGCGCGTCATCCGCCAGGCCCTCGCCAACGCGGACCTGACCCCCGCCGACGTC 20913
Db 9344 GCCCAGCAACGCGTCATCGCCGAGGCGCTCGCCGACGCCGGGCTGTCCCCGGAGGACGTG 9403
Qy 20914 GATGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATCGAGGCCCAGGCC 20973
Db 9404 GACGCGGTCGAGGCGCACGGCACCGGCACCCGGCTCGGCGACCCCATCGAGGCCGGGGCG 9463
Qy 20974 ATCCTCGCGACCTACGGACAGGACCGTCCCGGCAACGGGCCGTTGTGGCTGGGCTCCGTC 21033
Db 9464 CTGCTCGCCGCCTCCGGACGGAACCGTTCCGGCGACCACCCGCTGTGGCTCGGCTCGCTG 9523
Qy 21034 AAGTCCAACGTCGGACACACACAGGCCGCGGCGGGCGTGGCCGGAGTGATCAAGATGGTG 21093
Db 9524 AAGTCCAACATCGGGCATGCCCAGGCCGCCGCCGGTGTCGGCGGCGTCATCAAGATGCTC 9583
Qy 21094 ATGGCCCTCCGCCACCGGACACTCCCACCGACTCTCCACGCGGATGAGCCGTCGCCGCAT 21153
Db 9584 CAGGCGCTGCGGCACGGCTTGCTGCCCCGCACCCTCCACGCCGACGAGCCGACCCCGCAT 9643
Qy 21154 GTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCCGGCGGG 21213
Db 9644 GCCGACTGGAGCTCCGGCCGGGTACGGCTGCTCACCTCCGAGGTGCCGTGGC---AGCGG 9700
Qy 21214 GAGGGGCGGCCGCGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAACGCCCAC 21273
Db 9701 ACCGGCCGGCCCCGGCGGACCGGGGTGTCCGCCTTCGGCGTCGGCGGCACCAATGCCCAT 9760
Qy 21274 GTCATCCTCGAAGAAGCACCCGCCGACGACGTTCCGGGGGGACCACCCGCCGACGAGGAT 21333
Db 9761 GTCGTCCTCGAAGAGGCACCCGCC---------CCGCCCGCGCCGGAACCGGCCGGGGAG 9811
Qy 21334 GCCGGTAGTGGCGAGGAGGCTGCTGCCGGCAGTCCTGGGGTGTGGCCGTGGCTGGTGTCG 21393
Db 9812 GCCCCCGGCGGCTCCCGCGCCGCAGAAGGGGCGGAAGGGCCCCTGGCCTGGGTGGTCTCC 9871
Qy 21394 GCCAAGTCGCAGCCGGCCCTGCGCGCCCAGGCCCAGGCCCTGCACGCCCACCTCACCGAC 21453
Db 9872 GGACGCGACGAGCCGGCCCTGCGGTCCCAGGCCCGGCGGCTCCGCGACCACCTCTCCCGC 9931
Qy 21454 CACCCCGGCCTCGACCTCGCCGACGTCGGATACACCCTCGCCCACGCCCGCGCCGTGTTC 21513
Db 9932 ACCCCCGGGGCCCGCCCGCGTGACATCGCCTTCTCCCTCGCCGCCACGCGCGCAGCCTTT 9991
Qy 21514 GACCACCGCGCCACCCTCATCGCCGCCGACCGCGACACCTTCCTGCAAGCACTCCAGGCA 21573
Db 9992 GACCACCGCGCCGTGCTGATCGGCTCGGACGGGGCCGAACTCGCCGCCGCCCTGGACGCG 10051
Qy 21574 CTCGCCGCAGGCGAACCCCACCCCGCCGTCATCCACAGCAGCGCCCCAGGCGGGACCGGG 21633
Db 10052 TTGGCCGAAGG--------ACGCGACGGTCCGGCGGTGGTGCGCGGAGTCCGCGACCGGG 10103
Qy 21634 ACCGGGGAGGCCGCAGGAAAGACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCC 21693
Db 10104 AC-------------GGCAGGATGGCCTTCCTCTTCACCGGGCAGGGCAGCCAGCGCGCC 10150
Qy 21694 GGCATGGCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATC 21753
Db 10151 GGGATGGCCCACGACCTGCATGCCGCCCATACCTTCTTCGCGTCCGCCCTCGACGAGGTG 10210
Qy 21754 TGCACCCACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGAC 21813
Db 10211 ACGGACCGTCTCGACCCGCTGCTCGGCCGGCCGCTCGGCGCGCTGCT----------G 10258
Qy 21814 AACGACAACGACAACGAGGACGCGGCCGCACTGCTCCAGCAGACCCCGTACGCCCAGCCC 21873
Db 10259 GACGCCCGACCCGGCTCGCCCGAAGCGGCACTCCTGGACCGGACCGAGTACACCCAGCCG 10318
Qy 21874 GCCCTCTTCGCCTTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACC 21933
Db 10319 GCGCTCTTCGCCGTCGAGGTGGCGCTCCACCGGCTGCTG---GAGCACTGGGGGATGCGC 10375
Qy 21934 CCCCACTACTACGCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATC 21993
Db 10376 CCCGACCTGCTGCTGGGGCACTCGGTGGGCGAACTGGCGGCCGCCCACGTCGCGGGTGTG 10435
Qy 21994 CTCACCCTCACCGACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATG 22053
Db 10436 CTCGATCTCGACGACGCCTGCGCGCTGGTGGCCGCCCGCGGCAGGCTGATGCAGCGCCTG 10495
Qy 22054 CCCCC---CGGCACCATGACCACCCTCCACACCACCCCACACCACATCACCCACCACCTC 22110
Db 10496 CCGCCCGGCGGCGCGATGGTCTCCGTGCGGGCCGGCGAGGACGAGGTCCGCGCACTGCTG 10555
Qy 22111 ACCGCCCACGAAAACGACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATC 22170
Db 10556 GCCGGCCGCGAGGACGCCGTCTGCGTCGCCGCGGTGAACGGCCCCCGGTCGGTGGTGATC 10615
Qy 22171 AGCGGCACCCCCCACACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAA 22230
Db 10616 TCCGGCGCGGAGGAAGCGGTGGCCGAGGCGGCGGCGCAGCTCGCCGGACGAGGCCGCCGC 10675
Qy 22231 ACCAAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAAC 22290
Db 10676 ACCAGGCGGCTCCGCGTCGCGCACGCCTTCCACTCACCCCTGATGGACGGCATGCTCGCC 10735
Qy 22291 CAACTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACC 22350
Db 10736 GGATTCCGGGAGGTCGCCGCCGGCCTGCGCTACCGGGAACCGGAGCTGACGGTCGTCTCC 10795
Qy 22351 GCCAACA----------CCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAA 22398
Db 10796 ACGGTCACGGGGCGGCCCGCCCGCCCCGGTGAACTCACCGGCCCCGACTACTGGGTGGCC 10855
Qy 22399 CAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTC 22458
Db 10856 CAGGTCCGTGAGCCCGTGCGCTTCGCGGACGCGGTCCGCACGGCACACCGCCTCGGAGCC 10915
Qy 22459 ACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTC 22518
Db 10916 CGCACCTTCCTGGAGACCGGCCCGGACGGCGTGCTGTGCGGCATGGCAGAGGAGTGCCTG 10975
Qy 22519 CCCAACACCCCCACCACCACCCTCACCCTCACCCACCCCCACCACCACCCCCAAACCCAC 22578
Db 10976 GAGGACGACACCGTGGCCCTGCTGCCGGCGATCCACAAGCCCGGCACCGCG--------- 11026
Qy 22579 CTCCTCACCAACCTCGCCAAAACCACCACCACCTGGCACCCCCACCACTACACCCACCAC 22638
Db 11027 ---------- 11026
Qy 22639 CACAACCAACCCCACACCCACACCCACCTCGACCTCCCCACCTACCCCTTCCAACACCAC 22698
Db 11027 ---------- 11026
Qy 22699 CACTACTGGCTCGAACTACCCAGCGCCCAAACCAGCCCCGGTCAAAGGCGTTCTCGCCGC 22758
Db 11027 ----------CCGCACGGTCCGGCGGCTCCCGGCGCGCTGCGGGCGGCCGCC 11068
Qy 22759 TCGGCTCCAGACACCGCCGAGTCGGAGTTCTGGGACGCGGTGAACGAGGAAGACCTCCAG 22818
Db 11069 GCCGCGTACGGCCGGGGCGCCCGGGTGGACTGGGCCGGGATGCACGCCGACGGCCCCGAG 11128
Qy 22819 AGCCTCGCCGAAACCCTCGACATCGACGCCTCTGCTCTGGACACGGTGGTGCCCGCACTC 22878
Db 11129 GGGCCGGCCCGCCGCGTCGAACTGCCCGTCCACGCCTT----------CCGGCAC 11173


Qy 18647 ----CCCAAGCCCAAACCTGGGGACTCGCCCGCACCACCCTCCTCGAACACCCCACCCAC 18702
Db 7004 GAGCCGCGCTTCGACGGCTCGCGACTGGTGTGCGTCACCAGGGGCGCGGTCGCCGCACTC 7063
Qy 18703 ACCGCCGGAATCATCGACCTCCCCACCACCCCCACCCCCCACACCCTCCACCACCTCACC 18762
Db 7064 CCCGGCGAGGACGTCACCTCCCTCGCCACCGGCCCCCTCTGGGGCCTCGTCCGCTCCGCC 7123
Qy 18763 CAAACCCTCACCCAACCCCACC---------- 18784
Db 7124 CAGTCCGAGAACCCGGGACGCCTGTTCCTCCTGGACCTGGGTGAAGGCGAAGGCGAGCGC 7183
Qy 18785 ----------ACCAAACCCAACTCGCCATC 18804
Db 7184 GACGGAGCCGAGGAGCTGATCCGCGCGGCCACGGCCGGGGACGAGCCGCAGCTCGCGGCA 7243
Qy 18805 CGCACCACCGGCACCCACACCCGCCGCCTCACCCCCACCACCCTCACCCCCACACACCAA 18864
Db 7244 CGGGACGGCCGACTGCTCGCGCCGAGGCTGGCCCGTACCGCCGCCCTTTCGAGTGAGGAC 7303
Qy 18865 CCACCCACCCCCAC----------CCCCCACGGAACCACCCTCATCACCGGCGGAACC 18912
Db 7304 ACCGCCGGCGGCGCCGACCGTTTCGGCCCCGACGGCACCGTCCTCGTCACCGGGGGCACC 7363
Qy 18913 GGCGCCCTCGCCACCCACCTCACCCACCACCTCACCACCCACCAACCCACCCAACACCTC 18972
Db 7364 GGAGGCCTCGGAGCGCTCCTCGCCCGCCACCTCGTGGAGCGTCACGGGGTGCGCCGGCTG 7423
Qy 18973 CTCCTCACCAGCCGAACCGGCCCCCACACCCCCCACGCACAACACCTCACCACCCAACTC 19032
Db 7424 CTGCTGGTGAGCCGCCGCGGGGCCGACGCCCCCGGCGCGGCCGACCTGGGCGAGGACCTC 7483
Qy 19033 CAACAAAAAGGCATCCACCTCACCATCACCACCTGCGACACCAGCAACCCAGACCAACTC 19092
Db 7484 GCGGGCCTCGGCGCGGAGGTGGCGTTCGCCGCCGCCGACGCCGCCGACCGCGAGAGCCTG 7543
Qy 19093 CAACAACTCCTCAACACCATCCCCCCACAACACCCCCTCACCACCGTCATCCACACCGCA 19152
Db 7544 GCGCGGGCGATCGCCACCGTGCCCGCCGAGCATCCGCTGACGGCCGTCGTGCACACGGCG 7603
Qy 19153 GGCATCCTCGACGACGCCACCCTCACCAACCTCACCCCCACCCAACTCAACAACGTCCTC 19212
Db 7604 GGAGTCGTCGACGACGCGACGGTGGAGGCGCTCACACCGGAACGGCTGGACGCGGTACTG 7663
Qy 19213 CGCGCCAAAGCCCACAGCGCCCACCTCCTCCACCAACTCACCCAACACACCCCCCTCAAC 19272
Db 7664 CGCCCGAAGGTCGACGCCGCGTGGAACCTGCACGAGCTCACCAAGGACCTGCGGCTCGAC 7723
Qy 19273 GCCTTCGTCCTCTACTCCTCCGCCGCCGCCACCTTCGGCGCACCCGGCCAAGCCAACTAC 19332
Db 7724 GCCTTCGTCCTCTTCTCCTCCGTCTCCGGCATCGTCGGCACCGCCGGCCAGGCCAACTAC 7783
Qy 19333 GCCGCAGCCAACGCCTACCTCGACGCCCTCGCCCACCACCGCCACACCCACCACCTCCCC 19392
Db 7784 GCGGCGGCCAACACGGGCCTCGACGCCCTCGCCGCCCACCGCGCCGCCACGGGCCTGGCC 7843
Qy 19393 GCCACCAGCATCGCCTGGGGCACCTGGCAAGGAAACGGACTGGCGACTGGTCAAGTCAGC 19452
Db 7844 GCCACGTCGCTGGCCTGGGGCCTCTGGGACGGCACGCACGGCATGGGCGGCACGCTCGGC 7903
Qy 19453 GAACATCTCCGCCGCCGCGGGAT---------GTTCGCCATGCCGCCCGAGTTGGCGGTC 19503
Db 7904 GCCGCCGACCTCGCCCGCTGGAGCCGGGCCGGAATCACCCCGCTCACCCCGCTGCAGGGC 7963
Qy 19504 ACAGCTGTTGACGGCGCGATCGCGAGCGGGCGCCCGAGTCTCCTCGTCGCCGATATCGAC 19563
Db 7964 CTCGCGCTCTTCGACGCCGCGGTCGCCAGGGACGACGCCCTCCTCGTACCCGCCGGGCTC 8023
Qy 19564 TGGAAGAAATTGGGACCGGTTCTCTCCAGCAAGTCGTCGGTCTTGCTCGAGGACCTTCCC 19623
Db 8024 CGTCCCACCGCCCACCGGGGCACGGACGGACAGCCTCCTGCGCTGTGGCGCGGCCTCGTC 8083
Qy 19624 CAGGCACAGGGAACTGAGGAGGCGCGCAGTACCGTTGAGCAGACGGAGAGCACAAACCTC 19683
Db 8084 CGGGCGCGCCCGCGCCGTGCCGCGCGGACGGCCGCCGAGGCGGCGGACACGACCGGCGGC 8143
Qy 19684 CGGCAACTC---CTCATGGGTCGGTCACGTTCCGAGCAGGAAGAAGAGCTGCTCAGCCTC 19740
Db 8144 TGGCTGAGCGGGCTCGCCGCACAGTCCCCCGAGGAGCGGCGCAGCACAGCCGTCACGCTC 8203
Qy 19741 GTCCGCATCCACTCCGCGGCAGTGCTCGGGCGCGACGACTCCGAGGCCATCCCGCCCGGT 19800
Db 8204 GTGACGGGTGTCGTCGCGGACGTCCTCGGGCACGCCGACTCCGCCGCGGTCGGGGCGGAG 8263
Qy 19801 CGGCTGTTCAGGGATCTAGGGTTCGACTCGCTTGCGGCGGTGGAGCTTCGCAACCACCTC 19860
Db 8264 CGGTCCTTCAAGGACCTCGGCTTCGACTCCCTGGCCGGGGTGGAGCTCCGCAACCGGCTG 8323
Qy 19861 GCAGCACAGACGGAGCTGGCTCTGCCGACGACTCTCGTCTTCGATTACCCCAGCCCCACC 19920
Db 8324 AACGCCGCCACCGGCCTGCGGCTCCCCGCGACCACGGTCTTCGACCATCCCTCGCCGGCC 8383
Qy 19921 AAGCTCGCCCAATTTCTGCTCTCCGAGATCGCGGAGTTCCAGCCCGACAACTCAACTCCG 19980
Db 8384 GCGCTCGCGTCCCATCTCCTCGCCCAGGTGCCCGGGTTGAAGGAGGGGACGGCGGCGACC 8443
Qy 19981 CTTCC----------GCGACCCCGGGCAGAGCTCGATGAGCCG 20013
Db 8444 GCGACCGTCGTGGCCGAGCGGGGCGCTTCCTTCGGTGACCGTGCGACCGACGACGATCCG 8503
Qy 20014 ATCGCCATCGTTGGCATGGCCTGTCGCTTCCCCGGCGGAGTGACCTCGGCGGACGACTTC 20073
Db 8504 ATCGCGATCGTGGGCATGGCATGCCGCTATCCGGGTGGTGTGTCGTCGCCGGAGGACCTG 8563
Qy 20074 TGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACCGACCGCGGCTGG 20133
Db 8564 TGGCGGCTGGTGGCCGAGGGGACGGACGCGATCAGCGAGTTCCCCGTCAACCGCGGCTGG 8623
Qy 20134 GACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGCTACACCCGAAAC 20193
Db 8624 GACCTGGAGAGCCTCTACGACCCGGATCCCGAGTCGAAGGGCACCACGTACTGCCGGGAG 8683
Qy 20194 GGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGCATCAGCCCCCGC 20253
Db 8684 GGCGGGTTCCTGGAAGGCGCCGGTGACTTCGACGCCGCCTTCTTCGGCATCTCGCCGCGC 8743
Qy 20254 GAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCCTGGGAAACCATC 20313
Db 8744 GAGGCCCTGGTGATGGACCCGCAGCAGCGGCTGCTGCTGGAGGTGTCCTGGGAGGCGCTG 8803
Qy 20314 GAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGAGTCTTCACCGGC 20373
Db 8804 GAACGCGCGGGCATCGACCCGTCCTCGCTGCGCGGCAGCCGCGGTGGTGTCTACGTGGGC 8863
Qy 20374 ACCAACGGACAGGACCACGCGGCACACATCCGTCAGGCCCCGAGCGGTACCGAGGGATTC 20433
Db 8864 GCCGCGCACGGCTCGTACGCCTCCGATCCCCGGCTGGTGCCCGAGGGCTCGGAGGGCTAT 8923
Qy 20434 GTCCTGACCGGGGCAGCCACCAGCATCGCCTCCGGCCGAATCTCCTACATCCTCGGGTTG 20493
Db 8924 CTGCTGACCGGCAGCGCCGACGCGGTGATGTCCGGCCGCATCTCCTACGCGCTCGGTCTC 8983
Qy 20494 GAAGGGCCTGCGGTCACCCTCGACACAGCGTGTTCCTCCTCGCTCGTCGCCCTGCACCTC 20553
Db 8984 GAAGGACCGTCCATGACGGTGGAGACGGCCTGCTCCTCCTCGCTGGTGGCGCTGCATCTG 9043
Qy 20554 GCCTGCCAGTCCCTCAGGTCCGGTGAATGCACCATGGCCTTGGCCGGCGGGGCCACGGTC 20613
Db 9044 GCGGTACGGGCGCTGCGGCACGGCGAGTGCGGGCTCGCGCTGGCGGGCGGGGTGGCGGTG 9103
Qy 20614 ATGACCACCCCGATCACCTTCACCGAATTCGCCCGCCAACGCGGACTCGCCCCCGACGGG 20673
Db 9104 ATGGCCGATCCGGCGGCGTTCGTGGAGTTCTCCCGGCAGAAGGGGCTGGCCGCCGACGGC 9163
Qy 20674 CGTTGCAAGGCGTTCTCGGCGGCGGCTGACGGTACCGGCTGGGGTGAGGGTGTGGGGATG 20733


Db 4826 GTGTTCGCCGCCGCCCTCGACACCGTGTACGCGGCCCTCGACCGTCACCTCGACCGGCCG 4885
Qy 17056 CTCCTCCCCCTCCTCACCCAGGACCCCAACACCCAGGACACCACCACCCTCGAAGAAGCG 17115
Db 4886 CTGCGGGAGATCGTCGCC----------GCC 4906
Qy 17116 GCCGCACTGCTCCAGCAGACCCCGTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGCC 17175
Db 4907 GGGGAGGAGCTGGACCTCACCGCGTACACCCAGCCCGCCCTCTTCGCCTTCGAGGTGGCG 4966
Qy 17176 CTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTCC 17235
Db 4967 CTGTTCCGCCTCCT---CGAACACCACGGCCTCGTCCCCGACCTGCTCACCGGCCACTCC 5023
Qy 17236 CTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCACC 17295
Db 5024 GTCGGCGAGATCGCCGCCGCGCACGTCGCCGGTGTCCTCTCCCTCGACGACGCCGCACGT 5083
Qy 17296 CTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCC---CCCGGCACCATGACCACC 17352
Db 5084 CTCGTCACCGCCCGCGGCCGGCTCATGCAGTCGGCCCGCGAGGGCGGCGCGATGATCGCC 5143
Qy 17353 CTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGCC 17412
Db 5144 GTGCAGGCGGGCGAGGCCGAGGTCGTCGAGTCCCTGAAGGGCTACGAGGGCAGGGTCGCC 5203
Qy 17413 ATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCAA 17472
Db 5204 GTCGCCGCCGTCAACGGACCCACCGCCGTGGTCGTCTCCGGCGACGCGGACGCCGCCGAG 5263
Qy 17473 CACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAAAAAC 17532
Db 5264 GAGATCCGCGCCGTATGGGCGGGACGCGGCCGGCGCACCCGCAGGCTGCGCGTCAGCCAC 5323
Qy 17533 GCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAACC 17592
Db 5324 GCCTTCCACTCCCCGCACATGGACGACGTCCTCGACGAGTTCCTCCGGGTCGCCGAGGGC 5383
Qy 17593 CTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACAC----------CCCA 17640
Db 5384 CTGACCTTCGAGGAGCCGCGGATCCCCGTCGTCTCCACGGTCACCGGCGCGCTCGTCACG 5443
Qy 17641 CCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTAC 17700
Db 5444 TCCGGCGAGCTCACCTCGCCCGCGTACTGGGTCGACCAGATCCGGCGGCCCGTGCGCTTC 5503
Qy 17701 GCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCC 17760
Db 5504 CTGGACGCCGTCCGCACCCTGGCCGCCCAGGACGCGACCGTCCTCGTCGAGATCGGCCCC 5563
Qy 17761 GACAACACCCTCACCACCCTCACCCACCACAACCTCCCCAACACCCCCACCACCAC---- 17816
Db 5564 GACGCCGTCCTCACGGCACTCGCCGAGGAGGCTCTCGCGCCCGGCACGGACGCCCCGGAC 5623
Qy 17817 --------CCTCACCCTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTCACCAAC 17868
Db 5624 GCCCGGGACGTCACGGTCGTCCCGCTGCTGCGCGCGGGGCGCCCCGAGCCCGAGACCCTC 5683
Qy 17869 CTCGCCAAAACCACCACCACCTGGCACCCCCACCACTACACCCACCACCACAACCAACCC 17928
Db 5684 GCCGCCGGTCTCGCGACCGCCCATGTCCACGGCGCACCCTTGGACCGGGCGTCGTTCTTC 5743
Qy 17929 CACACCCACACCCACCTCGACCTCCCCACCTACCCCTTCCAACACCAGCACTACTGGCTC 17988
Db 5744 CCGGACGGGCGCCGCACGGACCTGCCCACGTACGCCTTCCGGCGCGAGCACTACTGGCTG 5803
Qy 17989 GAAAGCACACAGCCGGGTGCCGGATCCGGTTCGGGTTCCGGTTCCGGGCGGGCAGGGACT 18048
Db 5804 ACGCCCGAGGCCCGTACGGACGCCCGCGCACTCGGCTTCGACCCGGCGCGGCACCCGCTG 5863
Qy 18049 GCGGGCGGGACGGCAGAGGTGGAGTCGCGGTTCTGGGACGCGGTGGCCCGCCAGGACCTG 18108
Db 5864 CTGACGACCACGGTCGAGGTCGCCGGCGGCGACGGCGTCCTGCTGACCGGCCGTCTCTCC 5923
Qy 18109 GAAACGGTCGCGACCACGCTCGCCGTGCCCCCCTCCGCCGGCCTGGACACGGTGGTGCCC 18168
Db 5924 CTGACCGACCAGCCCTGGCTGGCCGACCACATGGTCAACGGCGCCGTCCTGTTGCCGGCC 5983
Qy 18169 GCACTCTCC---------- 18177
Db 5984 ACCGCCTTCCTGGAGCTCGCCCTCGCGGCGGGCGACCACGTCGGGGCGGTCCGGGTGGAG 6043
Qy 18178 ----------GCCTGGCACCGCCACCAACACGACCAAGCC 18207
Db 6044 GAACTCACCCTCGAAGCGCCGCTCGTCCTGCCCGAGCGGGGCGCCGTCCGCATCCAGGTC 6103
Qy 18208 CGCATCAACACCTGGACCTACCAGGAAACCTGGAAACCCCTCACCCTCCCCACCACCCAC 18267
Db 6104 GGCGTGAGCGGCGACGGCGAGTCGCCGGCCGGGCGCACCTTCGGTGTGTACAGCACCCCC 6163
Qy 18268 CAACCCCACCAAACCTGGCTCATCGCCATCCCCGAA---------- 18303
Db 6164 GACTCCGGCGACACCGGTGACGACGCGCCCCGGGAGTGGACCCGCCATGTCTCCGGCGTA 6223
Qy 18304 ---------- 18303
Db 6224 CTCGGCGAAGGGGACCCGGCCACGGAGTCGGACCACCCCGGCACCGACGGGGACGGTTCA 6283
Qy 18304 ----------ACCCAGACCCACCACCCCCACATCACCAACATCCTC 18339
Db 6284 GCGGCCTGGCCGCCTGCGGCGGCGACCGCCACACCCCTCGACGGCGTCTACGACCGGCTC 6343
Qy 18340 ACCAACCTCCACCA---------- 18353
Db 6344 GCGGAGCTCGGCTACGGATACGGTCCGGCCTTCCAGGGCCTGACGGGGCTGTGGCGCGAC 6403
Qy 18354 ---------- 18353
Db 6404 GGCGCCGACACGCTCGCCGAGATCCGGCTGCCCGCGGCGCAGCACGAGAGCGCGGGGCTC 6463
Qy 18354 ---------- 18353
Db 6464 TTCGGCGTACACCCGGCGCTGCTCGACGCGGCGCTCCACCCGATCGTCCTGGAGGGCAAC 6523
Qy 18354 ----------CCACGGCATCACCCCCATCCCCCTCACCCTC 18384
Db 6524 TCAGCTGCCGGTGCCTGTGACGCCGATACCGACGCGACCGACCGGATCCGGCTGCCGTTC 6583
Qy 18385 AACCACACCCACACCAACCCCCAACACCTCCACCACACCCGACAACAAGCCCAAAACCAC 18444
Db 6584 GCGTGGGCGGGGGTGACCCTCCACGCCGAAGGGGCCACCGCGCTCCGCGTACGGATCACA 6643
Qy 18445 ACCACCGGACCCATCACCGGCCTGCTCTCCCTCCTCGCCCTCGACGAAACACCCCACCCC 18504
Db 6644 CCCACCGGCCCGGACACGGTCACGCTCCGCCTCACCGACACCACCGGTGCGCCCGTGGCC 6703
Qy 18505 ---------- 18504
Db 6704 ACCGTGGAGTCCCTGACCCTGCGCGCGGTGGCGAAGGACCGGCTGGGCACCACCGCCGGG 6763
Qy 18505 ---------- 18504
Db 6764 CGCGTCGACGACGCCCTGTTCACGGTCGTGTGGACGGAGACCGGCACACCGGAACCCGCA 6823
Qy 18505 ----------CACCACCCCCACACACCCACCGGCACCCTCCTC 18537
Db 6824 GGGCGCGGAGCCGTGGAGGTCGAGGAACTCGTCGACCTCGCCGGCCTCGGCGACCTCGTG 6883
Qy 18538 AACCTCACCCTCACCCAAACCCACACCCAAACCCACCCACCAACCCCCCTCTGGTACGCC 18597
Db 6884 GAGCTCGGCGCCGCGGACGTCGTCCTCCGGGCCGACCGCTGGACGCTCGACGGGGACCCG 6943
Qy 18598 ACCACCAACGCCACCACCACCCACCCCAACGACCCCCTCACACACCCCA---------- 18646
Db 6944 TCCGCCGCCGCGCGCACAGCCGTCCGGCGCACCCTCGCCATCGTCCAGGAGTTCCTGTCC 7003


Qy 14880 ATCAACCCCGCTGCGTGATGTCCTCGTGGGCAAGTCTCCGCAGGAGCGAGACGAAGAGCT 14939
Db 2720 GGCCCGGCCGCGGGTGTGACGTCGGGCGAGGGCGAGGGCGAGGGCGAGGGCGCGGGTGCG 2779
Qy 14940 GTTGCGCCTG----------GTGCGCACCCATGCGGCC 14967
Db 2780 GGTGGCGGTGATCGGCCGGCTCGCCACGAGACGACCGAGCGCGTGCGCGCACACGTCGCC 2839
Qy 14968 GCTGTGCTGGGCCATGCCACTCCCGAAGTGATCGTTCCGAACAAGGCCTTCAAAGAGCTG 15027
Db 2840 GCCGTCCTCGAGTACGACGACCCGACCCGCGTCGAACTCGGCCTCACCTTCAAGGAGCTG 2899
Qy 15028 GGTTTTGATTCTCTCGCCGCAATTCAGCTTCGTAATCGACTGCTTGCTGACGTTGACCTG 15087
Db 2900 GGCTTCGACTCCCTCATGTCCGTCGAGCTGCGGAACGCGCTCGTCGACGACACGGGACTG 2959
Qy 15088 CCGCTTCCGGCCACGCTGATCTTCGATTACCCCACTCCGATGGCGCTT---------- 15135
Db 2960 CGCCTGCCCAGCGGACTGCTCTTCGACCACCCGACGCCGCGCGCCCTCGCCGCCCACCTG 3019
Qy 15136 TGCCAGTTCCTCCGGGCGGCGATCGTCGGAGCGGACACAGGCACGACCACTCGTCTGCCG 15195
Db 3020 GGCGACCTGCTCACCGGCGGCAGCGGCGAGACCGGATCGGCCGACGGGATACCGCCCGCG 3079
Qy 15196 CTAACTGCGGTCCCCGCCGACGAGCCGATCGCCATCGTCGGCATGGCCTGTCGGTACCCC 15255
Db 3080 ACCCCGGCGGACACCACCGCCGAGCCCATCGCGATCATCGGCATGGCCTGCCGCTACCCC 3139
Qy 15256 GGTGATGTACGGACGGTCGATGATCTCTGGCAGGTGGTCAGTGGTGGCCATGACGCGATC 15315
Db 3140 GGCGGCGTCACCTCCCCCGAGGACCTGTGGCGGCTCGTCGCCGAGGGGCGCGACGCCGTC 3199
Qy 15316 GGCGGATTCCCGACGAACCGTGGGTGGGACCTCGACACGCTGTACAACCCGGACCCGGAC 15375
Db 3200 TCGGGGCTGCCCACCGACCGCGGCTGGGACGAGGAC---CTCTTCGACGCCGACCCCGAC 3256
Qy 15376 CACCACGGAACCAGCTACACCCGGAGCGGCGGATTCCTTTACGACGCAGGCAATTTCGAT 15435
Db 3257 CGCAGCGGCAAGAGCTCGGTCCGCGAGGGCGGATTCCTGCACGACGCCGCCCTGTTCGAC 3316
Qy 15436 CCCGACTTCTTCGGTATCAGTCCGCGTGAGGCACTGGCGATGGACCCGCAGCAGCGGCTG 15495
Db 3317 GCCGGCTTCTTCGGGATATCGCCCCGCGAGGCCCTCGGCATGGACCCGCAGCAGCGGCTG 3376
Qy 15496 CTGCTGGAAACAGCGTGGGAGAGCATCGAACACGCCTGCATCAACCCCGACAGCCTCCGT 15555
Db 3377 CTCCTGGAGACGGCATGGGAGGCCGTGGAGCGCGCAGGGCTCGACCCCGAAGGCCTCAAG 3436
Qy 15556 GGCACACCAACCGGCGTCTTCGCCGGGCTGACCTACCACGACTACGCCGCGCGCTTTCCC 15615
Db 3437 GGCAGCCGGACGGCCGTCTTCGTCGGCGCCACCGCCCTGGACTACGGCCCGCGCATGCAC 3496
Qy 15616 ACAGCTCCGGCAGGGTTCGAGGGGTATCTCGGGCACGGAAGCGCAGGCAGTATCGCCTCG 15675
Db 3497 GACGGCGCCGAGGGCGTCGAGGGCCACCTCCTGACCGGGACCACGCCCAGCGTGATGTCG 3556
Qy 15676 GGTCGTGTCGCCTACGCTCTCGGCCTGGAAGGTCCGGCCCTCACAGTCGACACTGCCTGC 15735
Db 3557 GGCCGCATCGCCTACCAGCTCGGCCTCACCGGTCCTGCGGTCACCGTCGACACGGCCTGC 3616
Qy 15736 TCTTCGTCCCTGGTCGCTCTGCACCTGGCCTGTCAGGCGCTGCGGTCCGGCGAGTGTTCC 15795
Db 3617 TCGTCCTCGCTCGTCGCGCTGCACCTGGCCGTCCGTTCGCTGCGGCAGGGCGAGTCGAGC 3676
Qy 15796 ATGGCCCTCGCGGGTGGCGTCACGGTGATGTCAACCCCGGCCGGGTTCGTGGAGTTTTCG 15855
Db 3677 CTCGCGCTCGCCGGCGGAGCGACCGTCATGTCGACACCGGGCATGTTCGTCGAGTTCTCG 3736
Qy 15856 CGGCAGCGGGGCCTGGCCGTGGACGGGCGGTGCAAGGCGTTCTCGGCAGCGGCTGACGGC 15915
Db 3737 CGGCAGCGCGGCCTCGCCGCCGACGGCCGCTCCAAGGCCTTCTCCGACTCCGCCGACGGC 3796
Qy 15916 ACCGGCTGGGGTGAGGGTGTCGGAATGCTGCTGGTGGAGCGGCTGTCGGACGCGCGGCGG 15975
Db 3797 ACCTCCTGGGCCGAGGGCGTCGGCCTCCTCGTCGTCGAGCGGCTCTCGGACGCCGAGCGC 3856
Qy 15976 CTCGGTCACCGAATCCTCGCGGTGGTGCGTGGCAGTGCGGTCAATCAGGACGGTGCGAGC 16035
Db 3857 AACGGCCACCCCGTGCTCGCCGTGATCCGGGGCAGCGCGGTCAACCAGGACGGCGCCTCC 3916
Qy 16036 AACGGGCTGACGGCGCCCAACGGGCCGTCCCAGGAGCGTGTCATCCGCCTGGCCCTGGCC 16095
Db 3917 AACGGGCTCACCGCCCCCAACGGCCCGTCCCAGCAGCGCGTCATCCGACAGGCCCTGGCC 3976
Qy 16096 AACGCGGACCTGACCCCCGCCGACGTCGATGCGGTGGAGGCCCACGGCACCGGCACCACT 16155
Db 3977 GACGCCGGGCTCACCCCGGCCGACGTCGACGCCGTCGAGGCGCACGGTACGGGTACCCGG 4036
Qy 16156 TTGGGCGACCCGATCGAGGCCCAGGCCCTCCTCGCCACCTACGGACAGGACCGCCCCGGC 16215
Db 4037 CTCGGCGACCCCATCGAGGCCGAGGCGATCCTCGGCACCTACGGCCGGGACCGGGGCGAG 4096
Qy 16216 AACGAACCGCTGTGGCTGGGCTCGATGAAGTCGAACATCGGCCACGCGCAGGCTGCCGCA 16275
Db 4097 GGCGCTCCGCTCCAGCTCGGCTCGCTGAAGTCGAACATCGGCCACGCGCAGGCCGCCGCG 4156
Qy 16276 GGTGTGGGCGGGGTCATCAAGATGGTGATGGCGCTGCGGAATGGTCTGCTGCCGCGGACG 16335
Db 4157 GGCGTGGGCGGGCTCATCAAGATGGTCCTCGCGATGCGCCACGGCGTCCTGCCCAGGACG 4216
Qy 16336 TTGCATGTGGATGAGCCGTCGCCGCATGTGGACTGGTCCGCGGGGGCGGTGCAGCTGCTG 16395
Db 4217 CTCCACGTGGACCGGCCCACCACCCGCGTCGACTGGGAGGCCGGCGGCGTCGAGCTCCTC 4276
Qy 16396 ACGGAGACGGTGCCCTGGCCCGGCGGGGAGGGGCGGCTGCGGCGGGCAGGAGTGTCATCG 16455
Db 4277 ACCGAGGAGCGGGAGTGGCCGGA---GACGGGCCGCCCGCGCCGCGCGGCGATCTCCTCC 4333
Qy 16456 TTCGGCGTCAGCGGCACCAACGCCCACGTCATCCTCGAAGAAGCACCCGCCCACAACATC 16515
Db 4334 TTCGGCATCAGCGGCACCAACGCCCACATCGTGGTCGAACAGGCCCCGGAAGCCGGGGAG 4393
Qy 16516 CCGTCAGACACACCCGCCGACGACGCCCCGGGAGAAGCAGCCGCCGACGATGTTCCGGGG 16575
Db 4394 GCGGCGGTCACCACCACCGCCCCGGAAGCAGGGGAAG----------CCGGG 4435
Qy 16576 GAAGCGGCCGGCGACGACGCCGGTACCGGCGGGGAAGCGACTGGTCCTGCTGCCGGCAGT 16635
Db 4436 GAAGCGGCGGACACCACCGCCACCACGACGCCGGCCGCGGTCGG------CGTCCCCGAA 4489
Qy 16636 CCAGGGGTGTGGCCGTGGCTGGTGTCGGCCAAGTCGCAGCCGGCCCTGCGCGCCCAGGCC 16695
Db 4490 CCCGTACGCGCCCCCGTCGTGGTCTCCGCGCGGGACGCCGCCGCCCTGCGCGCCCAGGCC 4549
Qy 16696 CAGGCCCTGCACGCCCACCTCACCGACCACCCCGGCCTCGACCTCGCCGACGTCGGGTAC 16755
Db 4550 GTTCGGCTGCGGACCTTCCTCGACGGCCGACCGGACGTCACCGTCGCCGACCTCGGACGC 4609
Qy 16756 ACCCTCGCCCACGCCCGCGCCGTGTTCGACCACCGCGCCACCCTCATCGCCGCCGACCGC 16815
Db 4610 TCGCTGGC---CGCCCGTACCGCCTTCGAGCACAAGGCCGCCCTCACCACCGCCACCAGG 4666
Qy 16816 GACACCTTCCTGCAAGCACTCCAGGCACTCGCCGCAGGCGAACCCCACCCCGCCGTCATC 16875
Db 4667 GACGAGCTGCTCGCCGGGCTCGACGCCCTCGGCCGCGGGGAGCAAGCCACGGGCCTGGTC 4726
Qy 16876 CACAGCAGCGCCCCAGGCGGGACCGGGACCGGGGAGGCCGCAGGAAAGACCGCATTCATC 16935
Db 4727 ----------ACCGGCGAACCGGCCAGGGCCGGACGCACGGCCTTCCTG 4765
Qy 16936 TGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCACACCCACCCC 16995
Db 4766 TTCACCGGCCAGGGAGCGCAGCGCGTCGCCATGGGCGAGGAACTGCGCGCCGCGCACCCC 4825
Qy 16996 GTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTCGACCACCCC 17055


Db 617 GTGAACCTCAACATCCTCGCGGAGAGCGCCGTGACGGAGGAGCGCTTCGGTGGACTCTCC 676
Qy 12736 GCAGACGGCCGGTGCAAGGCGTTTTCGGCGGCGGCGGACGGGACCGGCTGGTCCGAGGGT 12795
Db 677 CCGGACGGCACCGCCTACACCTTCGACGCGCGGGCCAACGGATTCGTCCGGGGCGAGGGC 736
Qy 12796 GTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCGCCGCAACGGTCACCGTGTCCTG 12855
Db 737 GGCGGAGTCGTCGTACTCAAGCCGCTCTCCCGCGCCCTCGCCGACGGCGACCGTGTCCAC 796
Qy 12856 GCCGTCGTCCGCGGCAGCGCCGTCAACCAGGACGGCGCAAGCAACGGCCTGACCGCACCC 12915
Db 797 GGCGTCATCCGCGCCAGCGCCGTCAACAACGACGGAGCCACCCCGGGTCTCACCGTGCCC 856
Qy 12916 AACGGTCGTTCACAAGTCAAGGTCATCCGCCAGGCTTTGGCCAACGCACACCTCTCCCCT 12975
Db 857 AGCAGGGCCGCCCAGGAGAAGGTGCTGCGCGAGGCGTACCGGAAGGCGGCCCTGGACCCG 916
Qy 12976 GCCGATGTCGATGCGGTGGAGGCCCACGGCACGGGGACCACCCTGGGCGACCCGATCGAG 13035
Db 917 TCCGCCGTCCAGTACGTCGAACTCCACGGCACCGGAACCCCCGTCGGCGACCCCATCGAG 976
Qy 13036 GCTCAAGCCCTCGTCGAAGCCTACGGTCAGGACCGCCCCAACGGCCGCCCCCTCTGGCTC 13095
Db 977 GCCGCCGCGCTCGGCGCCGTCCTCGGCTCGGCGCGCCCCGCGGACGAACCCCTGCTCGTC 1036
Qy 13096 GGAACCCTCAAGTCCAACATCGGGCACTCCATGGCCGCTGCGGGTGTGGGCGGGGTCATC 13155
Db 1037 GGCTCGGCCAAGACGAACGTCGGGCACCTCGAAGGCGCCGCCGGCATCGTCGGCCTCATC 1096
Qy 13156 AAGATGGTGATGGCGCTGCGGAATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCG 13215
Db 1097 AAGACGCTCCTCGCGCTCGGCCGGCGCCGGATCCCGGCGAGCCTCAACTTCCGTACGCCC 1156
Qy 13216 TCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGG 13275
Db 1157 CACCCGGACATCCCGCTCGACACCCTCGGGCTCGACGTGCCCGACGGCCTGCGGGAGTGG 1216
Qy 13276 CCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACC 13335
Db 1217 CCGCACCCGGAC---CGCGAACTCCTCGCCGGCGTCAGCTCGTTCGGCATGGGCGGCACC 1273
Qy 13336 AACGCCCACGTCATCCTCGAGGAAGCACCCGCCCACAACATCCCGTCAGACACACCCGCC 13395
Db 1274 AACGCCCACGTCGTCCTCAGCGAAGGCCCCGCCCAG---------- 1309
Qy 13396 GACGACGTCCCGGGAGAATCAGCCGCCGACGAGGATGCCGGTAGTGGCGATGAGGCTGCT 13455
Db 1310 ----------GGCGGCGAGCAGCCCGGCATCGATGAGGAGACCCCC 1345
Qy 13456 GCCGGCAGTCCAGGGGTGTGGCCGTGGCTGGTGTCGGCCAAGTCGCAGCCGGCCCTGCGC 13515
Db 1346 GTCGACAGCGGGGCCGCACTGCCCTTCGTCGTCACCGGCCGCGGCGGCGAGGCCCTGCGC 1405
Qy 13516 GCCCAGGCCCAGGCCCTGCACGCCCACCTCACCGACCACCCCGGCCTCGACCTCGCCGAC 13575
Db 1406 GCCCAGGCCCGGCGCCTGCACGAGGCCGTCGAAGCGGACCCGGAGCTCGCGCCCGCCGCA 1465
Qy 13576 GTCGGGTACACCCTCGCCCACGCCCGCGCCGTGTTCGACCACCGCGCCACCCTCATCGCC 13635
Db 1466 CTCGCCCGGTCGCTGGTCACCACCCGTACGGTCTTCACGCACCGGTCGGTCGTCCTCGCC 1525
Qy 13636 GCCGACCGCGACACCTTCCTGCAAGCACTCCAGGCACTCGCCGCAGGCGAACCCCACCCC 13695
Db 1526 CCGGACCGCGCCCGCCTCCTCGACGGCCTCGGCGCCCTCGCCGCCGGGACGCCCGCGCCC 1585
Qy 13696 GCCGTCATCCACAGCAGCGCCCCAGGCGGGACCGGGACCGGGGAGGCCGCAGGAAAGACC 13755
Db 1586 GGCGTGGTCACCGGCACCCCC----------GCCCCCGGGCGCCTC 1621
Qy 13756 GCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCAC 13815
Db 1622 GCCGTCCTGTTCAGCGGCCAGGGTGCCCAACGTACGGGCATGGGCATGGAGTTGTACGCC 1681


Qy 13816 ACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTC 13875
Db 1682 GCCCACCCCGCCTTCGCGACGGCCTTCGACGCCGTCGCCGCCGAACTGGACCCCCTCCTC 1741
Qy 13876 GACCACCCCCTCCTCCCCCTCCTCACCCAGGACCCCAACACCCAGGACACCACCACCCTC 13935
Db 1742 GACCGGCCCCT----------CGCCGAACTC 1762
Qy 13936 GAAGAAGCGGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTC 13995
Db 1763 GTCGCGGCGGGCGACACCCTCGACCGCACCGTCCACACACAGCCCGCGCTCTTCGCCGTG 1822
Qy 13996 CAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCC 14055
Db 1823 GAGGTCGCCCTCCACCGCCTCGT---CGAGTCCTGGGGCGTCACGCCCGACCTGCTCGCC 1879
Qy 14056 GGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGAC 14115
Db 1880 GGCCACTCCGTCGGCGAGATCAGCGCCGCCCACGTCGCCGGGGTCCTGTCGCTGCGCGAC 1939
Qy 14116 GCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCC---CCCGGCACC 14172
Db 1940 GCCGCCCGCCTCGTCGCGGCGCGCGGCCGCCTCATGCAGGCGCTCCCCGAGGGCGGCGCG 1999
Qy 14173 ATGACCACCCTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAAC 14232
Db 2000 ATGGTCGCGGTCGAGGCGAGCGAGGAGGAAGTGCTTCCGCACCTCGCGGGACGCGAGCGG 2059
Qy 14233 GACCTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCAC 14292
Db 2060 GAGCTCTCCCTCGCGGCCGTGAACGGCCCCCGCGCGGTCGTCCTCGCGGGCGCCGAGCGC 2119
Qy 14293 ACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCC 14352
Db 2120 GCCGTCCTCGACGTCGCCGAGCTGCTGCGCGAACAGGGCCGCCGGACGAAGCGGCTCAGC 2179
Qy 14353 ACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCAC 14412
Db 2180 GTCTCGCACGCCTTCCACTCGCCGCTCATGGAGCCGATGCTCGACGACTTCCGCCGGGTC 2239
Qy 14413 ACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACAC------- 14465
Db 2240 GTCGAAGAGCTGGACTTCCAGGAGCCCCGCGTCGACGTCGTGTCCACGGTGACGGGCCTG 2299
Qy 14466 -----CCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACC 14520
Db 2300 CCTGTCACAGCGGGCCAATGGACCGATCCCGAGTACTGGGTGGACCAGGTCCGCAGGCCC 2359
Qy 14521 GTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAA 14580
Db 2360 GTACGCTTCCTCGACGCCGTACGCACCCTGGAGGAATCGGGCGCCGACACCTTCCTGGAG 2419
Qy 14581 CTCGGACCCGACAACACCCTCACCACCCTCACCCACGACAACCTCCCCAACACCCCCACC 14640
Db 2420 CTCGGTCCCGACGGGGTCTGCTCCGCGATGGCGGCGGACTCCGTACGCGACCAGGAGGCC 2479
Qy 14641 ACCACCCTCACCCTCACCCACCCCCACCACCACCCCCAAACCCACCTCCTCACCAACCTC 14700
Db 2480 GCCACGGCGGTCTCCGCCCTGCGCAAGGGCCGCCCGGAGCCCCAGTCGCTGCTCGCCGCA 2539
Qy 14701 GCCAAAACCACCACCACCTGGCACCCCCACCACTACACCCACCACCACAACCAACCCCAC 14760
Db 2540 CTCACCACCGTCTTCGTCCGGGGCCACGACGTCGACTGGACCGCCGCGCACGGGAGCACC 2599
Qy 14761 ACCCACACCCACCTCGACCTCCCCACCTACCCCTTCCAACACCACCACTACTGGCTCCA- 14819
Db 2600 GGCACGGTCAGGGTGCCCCTGCCGACCTACGCCTTCCAGCGCGAACGCCACTGGTTCGAC 2659
Qy 14820 ACCACCCGGCAAGCCGAGCGACCCGTCACCGAGCGAAGGCCGTGAGCAAGCCACGACCCC 14879
Db 2660 GGCGCCGCGCGAACGGCGGCGCCGCTCACGGCGGGCCGATCGGGCACCGGTGCGGGCACC 2719

```
Qy      26093 ATACCGACGACGAACCGCCTGCCAACGCACCAGCCCTGCCCCATCCCCTCCCTCTTCCCG 26152
              |  | ||    | | | | ||      ||  | |           |   ||     | ||
Db      19660 AGCCGGAGCCGGGAACTCGTGTGGTTGCTGCCGG---------TGATCTGGTGGTGCCGT 19710

Qy      26153 TGCCGGTGTCGGCGAGGTCTGAGGCCGGGTTGCGGGCGCAGGCACAGGCGTTGCGCCAGT 26212
                | |||||| | | ||  || ||  | |||||| || |||||       |||      |
Db      19711 GGGTGGTGTCCGGGCGGGATGCGGGGGCGTTGCGTGCACAGGCGGCACGCTTGGCTGCGC 19770

Qy      26213 ACGTGGCAGCCCGCCCGGACATGTCACCTGCCGACATTGGTGCGGGTCTGGCCCGCGGCC 26272
              | ||| |     || |||    |     |  ||  | ||   |   |||    |
Db      19771 ATGTGTCGA---GCACGGGTGCGGGTGTGGTTGATGTGGGCTGGTCGTTGGTGGCCACGA 19827

Qy      26273 GGGCCGTACTGGAACACCGCGCCGTCATCCTGGCCGCGGACCGCGAGGAACTGGCGCAGG 26332
              || | ||  | || ||||| || || ||  | | | | || |  ||     |||||  |
Db      19828 GGTCGGTGTTCGAGCACCGGGCGGTGATGGTCGGCACTGATCTTGATTCCATGGCGGGGT 19887

Qy      26333 CACTGACAGCCCTGGCAGCCGGCGAACCCCACCCCCACATCACCACAGGCCACACCCGGG 26392
              |  || | |   | || || || |          ||      |    | ||     | | |
Db      19888 CGTTGGCCGGGTTTGCTGCGGGTGGGGTTGTGCCGGGGGTGGTGTCGGGTGTGGCTCCG- 19946

Qy      26393 GCGGTGACCGCGGCGGCGTCGTCTTCGTCTTCCCCGGACAGGGCGGCCAGTGGGCCGGGA 26452
                | |||  |  |  | || || |||||||| || || |||||    |||||||  ||||
Db      19947 --GCTGAGGGTCGTCGTGTGGTGTTCGTCTTTCCTGGTCAGGGTTCGCAGTGGGTGGGGA 20004

Qy      26453 TGGGCCTGACCCTGCTCACCTCCTCACCCGTGTTCGCCGAACACATCGACGCATGCGAGA 26512
              |||           ||||      | |  || |||||||| ||     | |  |  || |
Db      20005 TGGCGGCTGGGTTGCTGGATGCGTGTCCGGTGTTCGCGGAGGCGGTGGCGGAGTGTGCCG 20064

Qy      26513 AAGCCCTCACCCCCTGGGTGCCCTGGTCCCTGACCGACATCCTGCACCGCGACCCCGACG 26572
              |  ||   ||| | |       ||||| ||| |||  |  ||| | | |      |  |
Db      20065 CGGTGCTGGACCCGTTGACCGGTTGGTCGCTGGTCGAGGTGTTGCGCGGTGGTGAGGCTG 20124

Qy      26573 ACCCCGCATGGCAACAAGCCGACGTGGTCCAGCCCGTGCTCTTCAGCATCATGGTCTCCC 26632
              |  |  ||      |  || ||||| ||||| | | |  |      | ||||| || |
Db      20125 TTCTTGGGCGG------GTTGATGTGGTGCAGCCGGCGTTGTGGGCGGTGATGGTGTCAC 20178

Qy      26633 TCGCCGCCCTGTGGCGCTCCTACGGCATCGAACCCGACGCGGTCCTCGGCCACTCCCAGG 26692
              | ||          ||||| |  ||||| | || || |  |||||  | || || |||| 
Db      20179 TGGCTCGGACCTGGCGGTATTACGGTGTGGAGCCTGCTGCGGTTGTGGGGCATTCGCAGG 20238

Qy      26693 GAGAAATCGCCGCCGCCCACATCTGCGGCGCACTCAGCCTGAAAGACGCCGCCAAAACCG 26752
              | ||| || || || ||      |  || |   | || |||   || |  ||       |
Db      20239 GTGAGATTGCTGCGGCTTGTGTGGCTGGGGGGTTGAGTCTGGCCGATGGTGCGCGGGTGG 20298

Qy      26753 TTGCACTGCGCAGCCGCGCACTGGCCGCCGTACGAGGCCGGGGCGCCATGGCCTCACTGC 26812
              | |  |||| ||||| ||  | |||    |    ||  |||||| ||||| |||  |
Db      20299 TGGTGTTGCGGAGCCGGGCGATCGCCCGGATCGCCGGTGGGGGCGGCATGGTCTCCGTCA 20358

Qy      26813 CCCTGCCCGCCCAGGACGTGCAGCAGCTCATTTCCGAACGGTGGGAAGGGCAGTTGTGGG 26872
                || || |||   | |       |   |||  |||    |  | || | ||||| ||
Db      20359 GTCTCCCGGCC---GGCCGTGTCCGCACCATGCTCGACACCTACGGCGGCCGGTTGTCGG 20415

Qy      26873 TGGCAGCCCTCAACGGCCCCCACTCCACCACCGTCTCCGGCGACACCAAGGCGGTGGATG 26932
              |||| ||  |||||||||||    ||| |||    || ||||| ||| || ||||  ||||||
Db      20416 TGGCTGCGGTCAACGGCCCGTCCTCGACCGTGGTGTCCGGTGACGCCCAGGCCCTGGATG 20475

Qy      26933 AGGTGCTGGCGCACTGCACCGACACCGGCCTACGGGCCAAACGCATCCCCGTCGACTACG 26992
              || || ||||   |||          ||  | |||||    ||  |||| || ||||| |
Db      20476 AGTTGTTGGCCGGCTGTGAGCGGGAGGGGGTCCGGGCTCGTCGTGTCCCGGTGGACTATG 20535

Qy      26993 CCTCCCACTGCCCCCACGTCCAACCCCTCCACGACGAACTCCTGCACCTGCTGGGAGACA 27052
              |||||||| | | ||  |  | |  | ||| || || ||| |  |||  ||||  ||||
Db      20536 CCTCCCACTCCGCGCAGATGGACCAGTTACGCGATGAGCTGCTGGAAGCGCTGGCGGACA 20595

Qy      27053 TCACCCCCCAGCCGTCCACCGTGCCGTTCTTCTCCACCGTGGAAGGCACCTGGCTGGACA 27112
              |||| || || |  |||| ||| |||||||||||| || |||    |  ||||||||||||
Db      20596 TCACTCCGCAACACTCCAGCGTTCCGTTCTTCTCGACGGTGACGGCGGACTGGCTGGACA 20655

Qy      27113 CCACAACCCTGGACGCCGCCTACTGGTACCGCAACCTCCACCAGCCCGTCCGCTTCAGCC 27172
              | ||  | ||||| || |  |||||||  |   || || |   || | ||||| |||
Db      20656 CGACCGCTCTGGATGCGGGGTACTGGTTCACGAATCTGCGGGAGACGGTCCGGTTCCAGG 20715

Qy      27173 ACGCCATCCAGACCCTGACCGACGACGGACACCGCGCCTTCATCGAAATCAGCCCCCACC 27232
```

```
                |  ||| || |    ||    |   | ||     |||| ||| ||||   |||||| ||||
Db      20716 AAGCCGTCGAAGGGCTTGTGGCTCAGGGGATGGGCGCGTTCGTCGAGTGCAGCCCACACC 20775

Qy      27233 CCACCCTCGTCCCCCGCCATCGAAGACACCACCGAAAACACCACCGAAAACATCACCGCGA 27292
              ||  ||||||||||||  |||||  | |||  |||  | |       |   |     |||
Db      20776 CCGTCCTCGTCCCCGGTATCGAGCAGACCCTCGACACCGTGGAAGCCGATGCTGTGGCGC 20835

Qy      27293 CCGGCAGCCTCCGCCGCGGCGACAACGACACCCACCGCTTCCTCACCGCCCTCGCCCACA 27352
                ||    || || || |  ||   || |      || ||||||||  |||||||  |
Db      20836 TGGGTTCGCTACGGCGTGATGAGGGCGGCCTGGGACGGTTCCTCACGTCCCTCGCGGAAG 20895

Qy      27353 CCCACACCACCGGCATC 27369
              ||  |  |   ||| ||
Db      20896 CCTTCGTCCAGGGCGTC 20912
```

RESULT 11
US-09-860-846-1
; Sequence 1, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 15872
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-1

Query Match 7.1%; Score 2188.4; DB 9; Length 15872;
Best Local Similarity 52.5%; Pred. No. 0;
Matches 6973; Conservative 0; Mismatches 5061; Indels 1248; Gaps 50;

```
Qy      12256 CGCAGCGGAGGGTTTCTCTATGACGCGGGCGACTTCGACGCCGCGTTCTTCGAGTTGTCA 12315
              ||| | || || || ||| |    |  | |  |||||||| || |||||||    | ||
Db        197 CGCCGGGGCGGCTTCCTCGACTCCCTCGACCTCTTCGACGCGGCCTTCTTCGGAATCTCG 256

Qy      12316 CCGCGTGAGGCGCTGGCAATGGACCCGCAGCAGCGCCTGCTGCTCGAAACCACTTGGGAA 12375
              || ||||||||    || |||||||||||||||| ||| | ||||||  | | |||||
Db        257 CCCCGTGAGGCCGCCGCCATGGACCCGCAGCAGCGACTGGTCCTCGAACTCGCCTGGGAG 316

Qy      12376 ACGTTCGAACAGGGCGGAATCGACCCGAGGTCCATGCGCGGAAGCCGGACCGGGGTTTTC 12435
              || | ||  | | ||||||||| |||  |  || |   |||||||| ||||  || |||
Db        317 GCGCTGGAGGACGCCGGAATCGTCCCCGGCACCCTCGCCGGAAGCCGCACCGCCGTCTTC 376

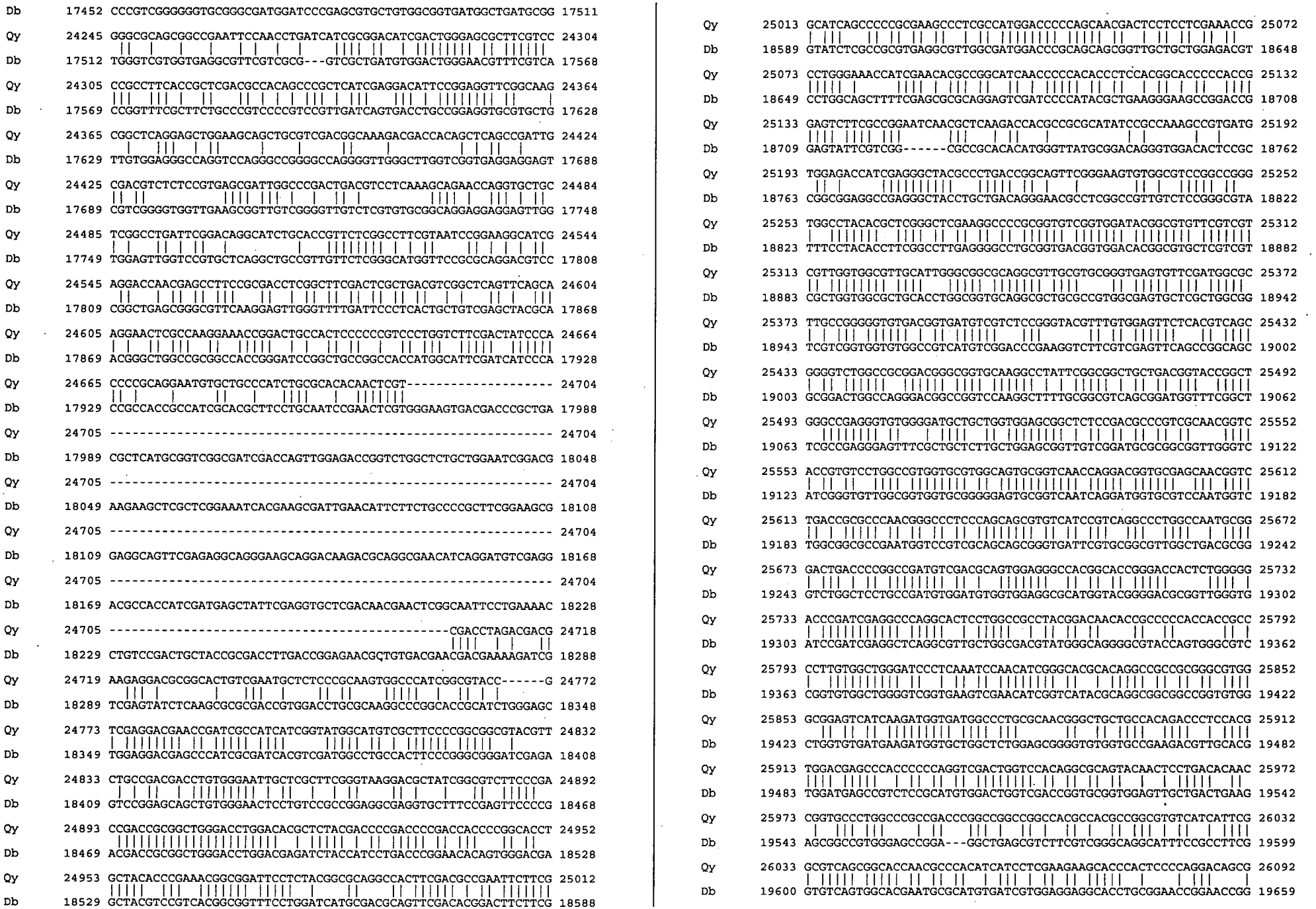
Qy      12436 GTGGGGATCAATCCGGAGGACTACACCACCGGATACACACATCAGCCCTCAAACGCAGTC 12495
              || || | |   | ||| |||||||| | |      |   || ||   |    | ||  ||
Db        377 GTCGGCACCCTGCGGGACGACTACACGAGCCTCCTCTACCAGCACGGCGAGCAGGCCATC 436

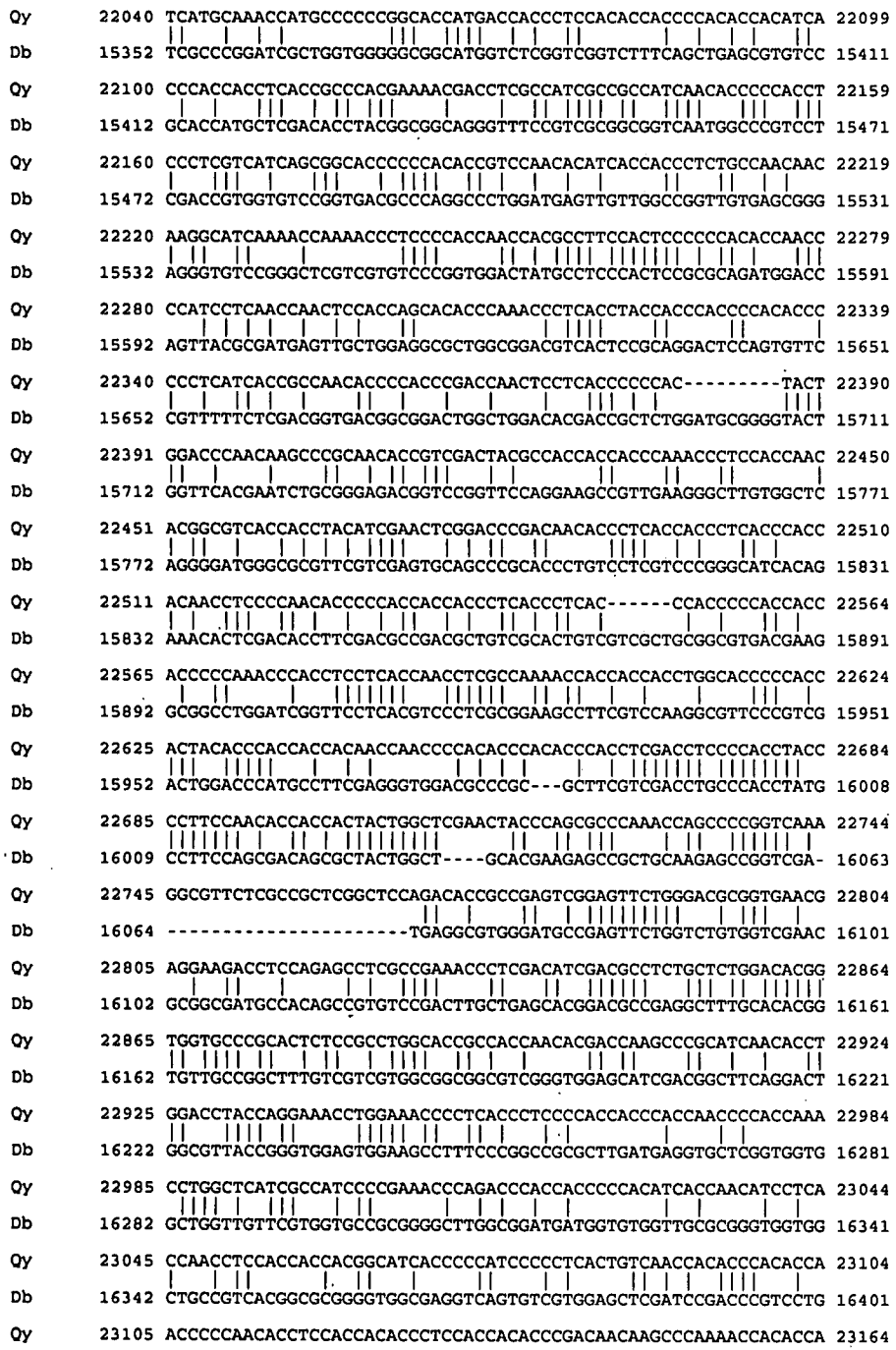
Qy      12496 GAGGGCTACCTGCTCACTGGCAGCGCGGCAAGCATTGCGTCAGGCCGTATCTCCTACAAC 12555
                ||    |  | |||              || |      |   |||  |||| ||| ||
Db        437 ACCCAGCACACCATGGCGGGCGTGAACCGGGGCGTCATCGCCAACCGCGTCTCGTACCAC 496

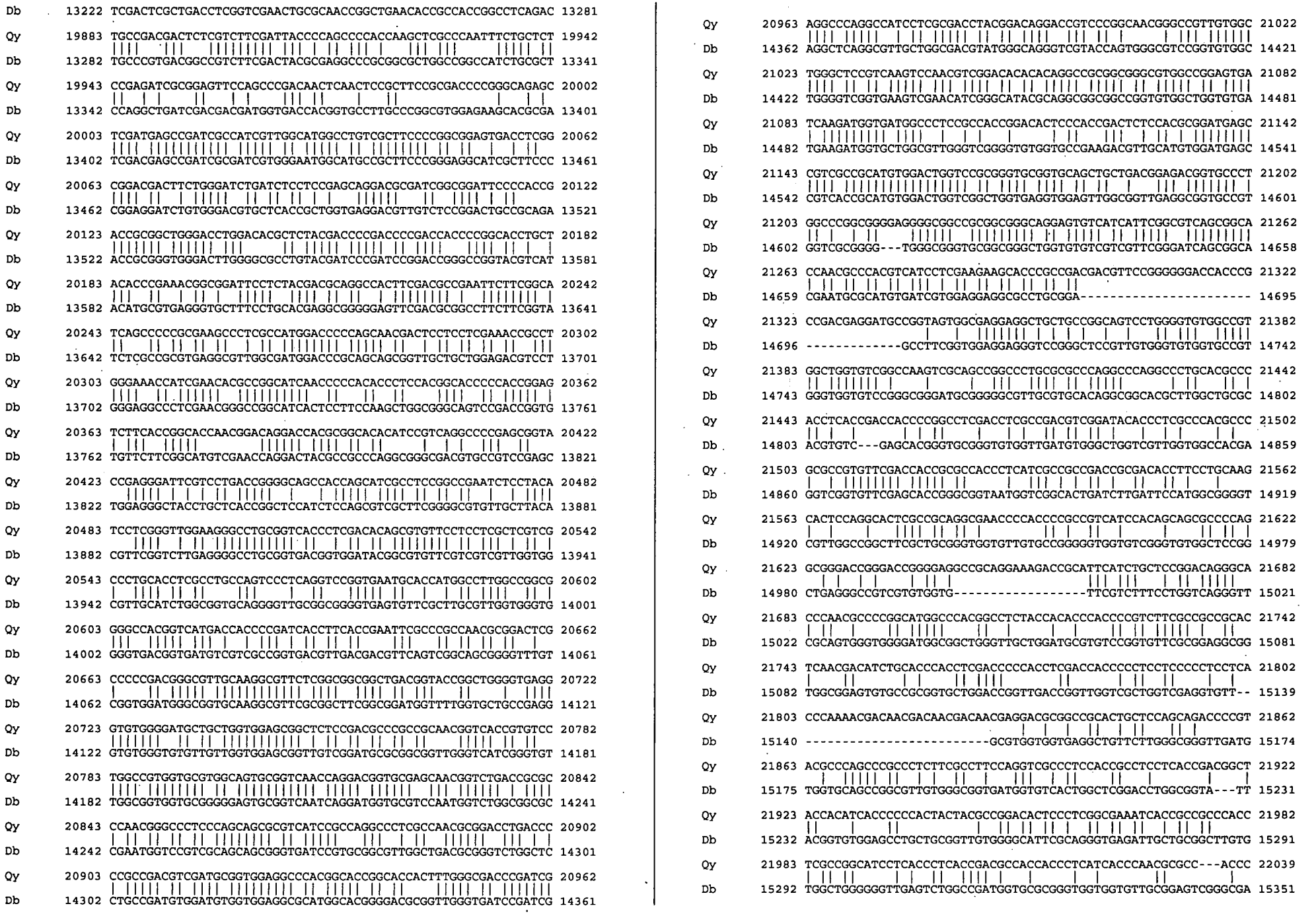
Qy      12556 TTCGGGCTCGAAGGCCCTGCGATCACTATCGACACCGCGTGTTCCTCCTCGCTCGTCGCC 12615
              |||| ||  | |||||      |||| |||||  |||||   || ||||||||||||||
Db        497 CTCGGCCTGCAGGGCCCGAGCCTCACCGTCGACGCCGCGCAGTCGTCCTCGCTCGTCGCC 556

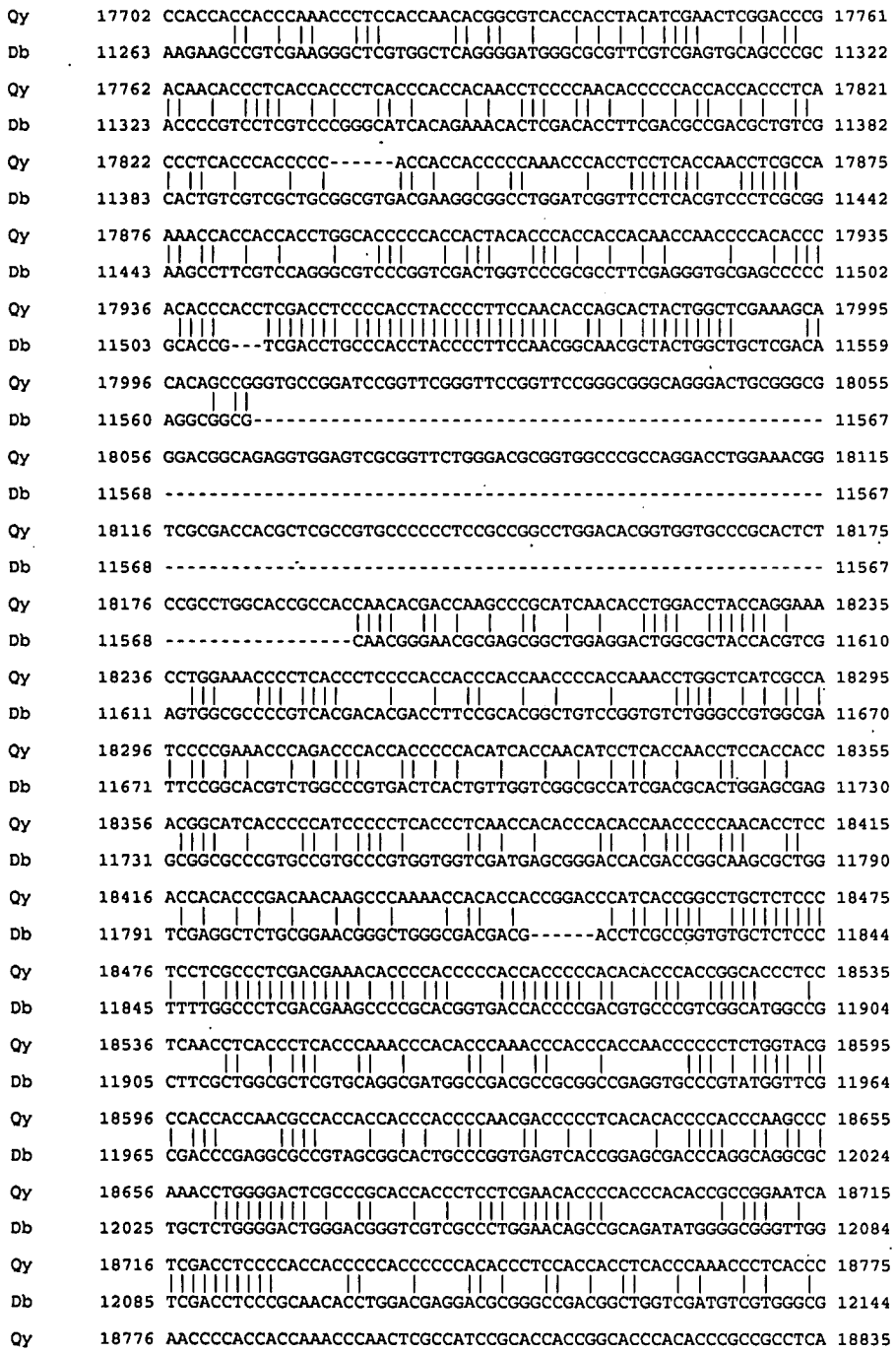
Qy      12616 CTGCATCTGGCCTGCCAAGCGCTCCGGTCCGGTGAATGCACCATGGCGCTCGCAGGCGGC 12675
              |||| |||||||||  |  | || ||  |||| || | ||| | ||||||||  | ||||
Db        557 GTGCACCTGGCCTGCGAGTCCCTGCGCGCCGGGGAGTCCACGACGGCGCTCGTCGCCGGC 616

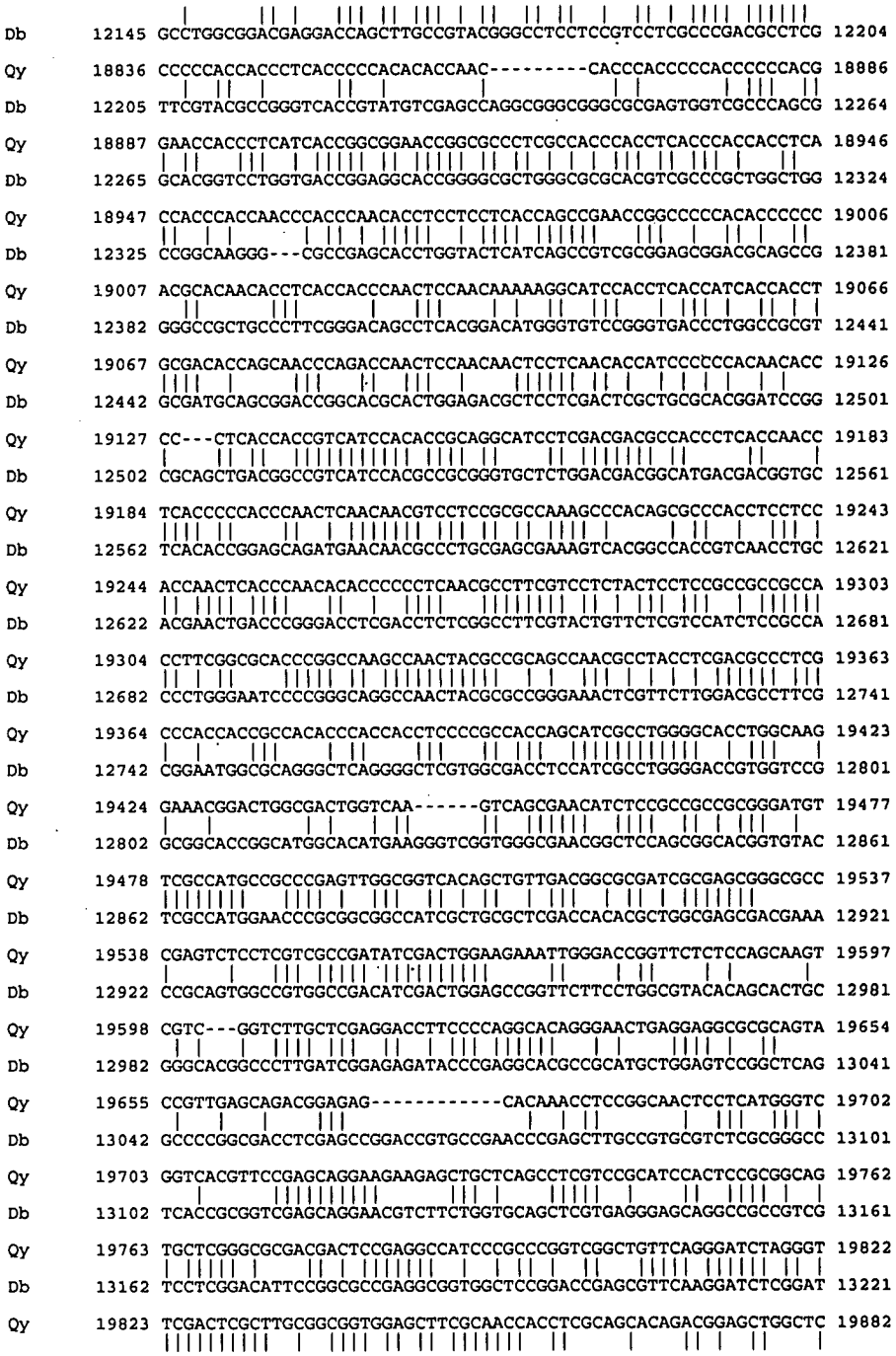
Qy      12676 GCCTCCGTCATGGCCACTCCCTTCGTCTTCACCGAGTTCTCTCGCCAGCGGGGCCTGGCC 12735
              |    | |||    |   |     |  |     |      |||    | || ||  ||
```

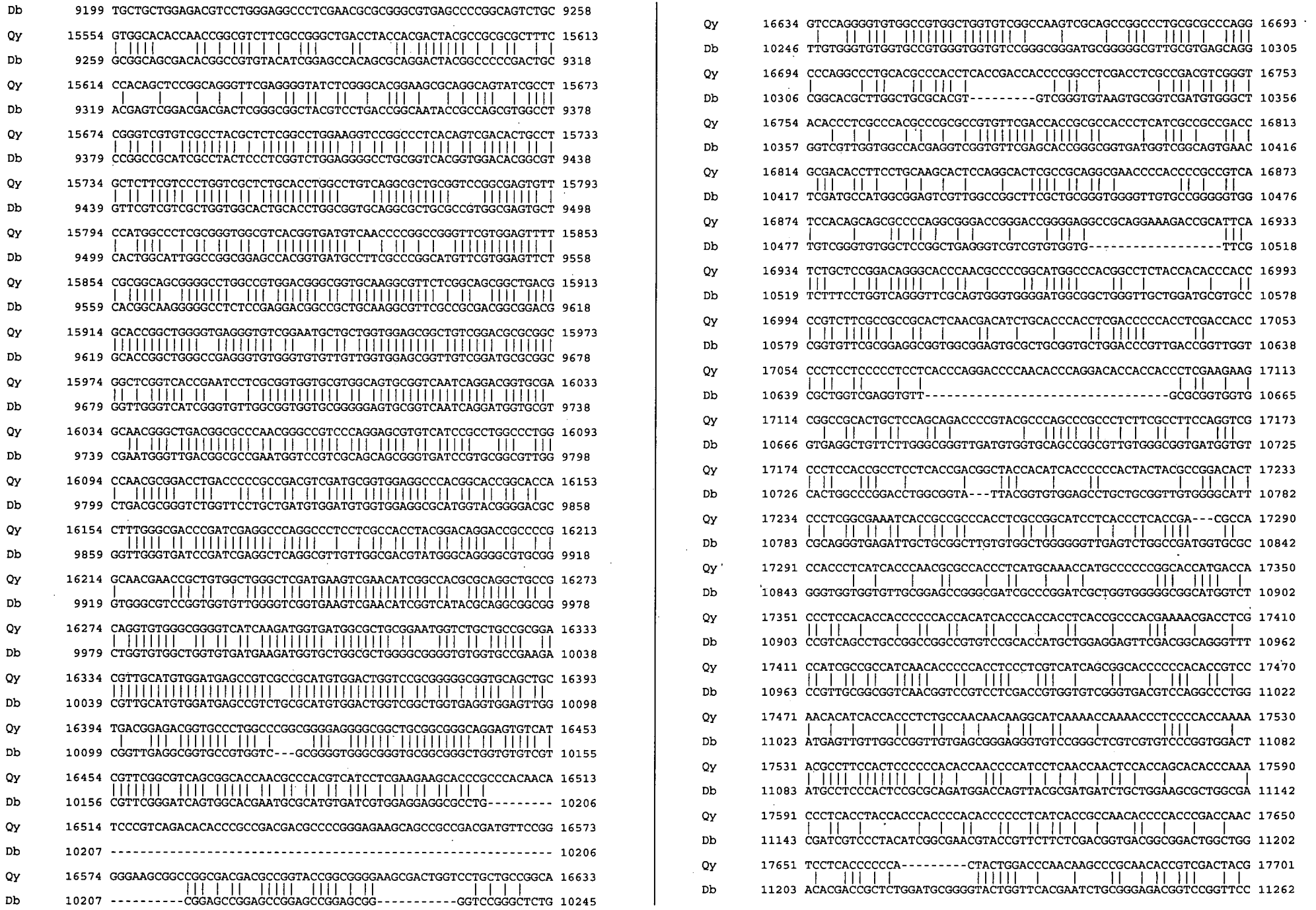


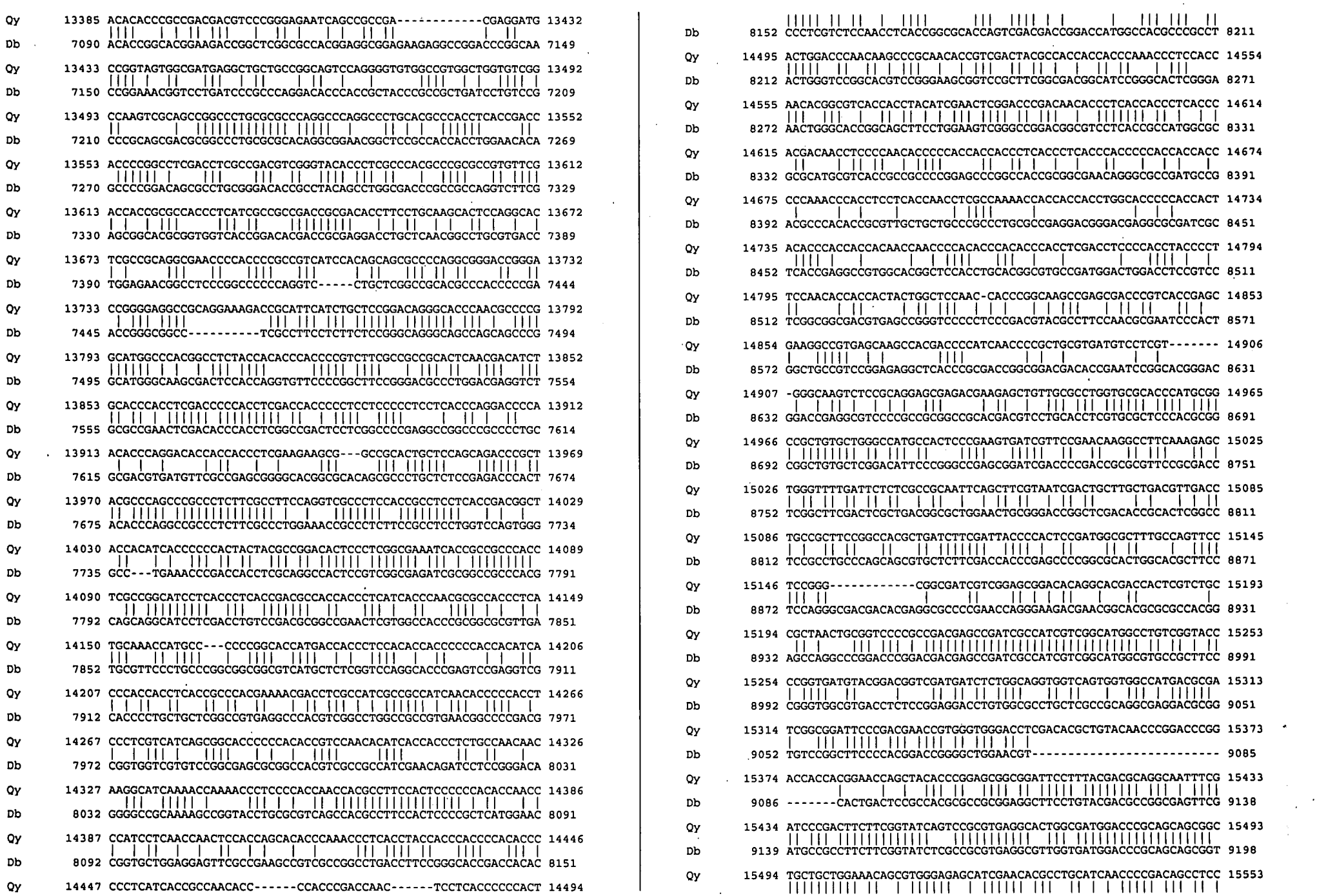

Qy 22040 TCATGCAAACCATGCCCCCCGGCACCATGACCACCCTCCACACCACCCCACACCACATCA 22099
Db 15352 TCGCCCGGATCGCTGGTGGGGGCGGCATGGTCTCGGTCGGTCTTTCAGCTGAGCGTGTCC 15411
Qy 22100 CCCACCACCTCACCGCCCACGAAAACGACCTCGCCATCGCCGCCATCAACACCCCCACCT 22159
Db 15412 GCACCATGCTCGACACCTACGGCGGCAGGGTTTCCGTCGCGGCGGTCAATGGCCCGTCCT 15471
Qy 22160 CCCTCGTCATCAGCGGCACCCCCCACACCGTCCAACACATCACCACCCTCTGCCAACAAC 22219
Db 15472 CGACCGTGGTGTCCGGTGACGCCCAGGCCCTGGATGAGTTGTTGGCCGGTTGTGAGCGGG 15531
Qy 22220 AAGGCATCAAAACCAAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACC 22279
Db 15532 AGGGTGTCCGGGCTCGTCGTGTCCCGGTGGACTATGCCTCCCACTCCGCGCAGATGGACC 15591
Qy 22280 CCATCCTCAACCAACTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCACACCC 22339
Db 15592 AGTTACGCGATGAGTTGCTGGAGGCGCTGGCGGACGTCACTCCGCAGGACTCCAGTGTTC 15651
Qy 22340 CCCTCATCACCGCCAACACCCCACCCGACCAACTCCTCACCCCCCAC---------TACT 22390
Db 15652 CGTTTTTCTCGACGGTGACGGCGGACTGGCTGGACACGACCGCTCTGGATGCGGGGTACT 15711
Qy 22391 GGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCACCCAAACCCTCCACCAAC 22450
Db 15712 GGTTCACGAATCTGCGGGAGACGGTCCGGTTCCAGGAAGCCGTTGAAGGGCTTGTGGCTC 15771
Qy 22451 ACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCACC 22510
Db 15772 AGGGGATGGGCGCGTTCGTCGAGTGCAGCCCGCACCCTGTCCTCGTCCCGGGCATCACAG 15831
Qy 22511 ACAACCTCCCCAACACCCCCACCACCACCCTCACCCTCAC------CCACCCCCACCACC 22564
Db 15832 AAACACTCGACACCTTCGACGCCGACGCTGTCGCACTGTCGTCGCTGCGGCGTGACGAAG 15891
Qy 22565 ACCCCCAAACCCACCTCCTCACCAACCTCGCCAAAACCACCACCACCTGGCACCCCCACC 22624
Db 15892 GCGGCCTGGATCGGTTCCTCACGTCCCTCGCGGAAGCCTTCGTCCAAGGCGTTCCCGTCG 15951
Qy 22625 ACTACACCCACCACCACAACCAACCCCACACCCACACCCACCTCGACCTCCCCACCTACC 22684
Db 15952 ACTGGACCCATGCCTTCGAGGGTGGACGCCCGC---GCTTCGTCGACCTGCCCACCTATG 16008
Qy 22685 CCTTCCAACACCACCACTACTGGCTCGAACTACCCAGCGCCCAAACCAGCCCCGGTCAAA 22744
Db 16009 CCTTCCAGCGACAGCGCTACTGGCT----GCACGAAGAGCCGCTGCAAGAGCCGGTCGA- 16063
Qy 22745 GGCGTTCTCGCCGCTCGGCTCCAGACACCGCCGAGTCGGAGTTCTGGGACGCGGTGAACG 22804
Db 16064 ----------TGAGGCGTGGGATGCCGAGTTCTGGTCTGTGGTCGAAC 16101
Qy 22805 AGGAAGACCTCCAGAGCCTCGCCGAAACCCTCGACATCGACGCCTCTGCTCTGGACACGG 22864
Db 16102 GCGGCGATGCCACAGCCGTGTCCGACTTGCTGAGCACGGACGCCGAGGCTTTGCACACGG 16161
Qy 22865 TGGTGCCCGCACTCTCCGCCTGGCACCGCCACCAACACGACCAAGCCCGCATCAACACCT 22924
Db 16162 TGTTGCCGGCTTTGTCGTCGTGGCGGCGGCGTCGGGTGGAGCATCGACGGCTTCAGGACT 16221
Qy 22925 GGACCTACCAGGAAACCTGGAAACCCCTCACCCTCCCCACCACCCACCAACCCCACCAAA 22984
Db 16222 GGCGTTACCGGGTGGAGTGGAAGCCTTTCCCGGCCGCGCTTGATGAGGTGCTCGGTGGTG 16281
Qy 22985 CCTGGCTCATCGCCATCCCCGAAACCCAGACCCACCACCCCCACATCACCAACATCCTCA 23044
Db 16282 GCTGGTTGTTCGTGGTGCCGCGGGGCTTGGCGGATGATGGTGTGGTTGCGCGGGTGGTGG 16341
Qy 23045 CCAACCTCCACCACCACGGCATCACCCCCATCCCCCTCACTGTCAACCACACCCACACCA 23104
Db 16342 CTGCCGTCACGGCGCGGGGTGGCGAGGTCAGTGTCGTGGAGCTCGATCCGACCCGTCCTG 16401
Qy 23105 ACCCCCAACACCTCCACCACACCCTCCACCACACCCGACAACAAGCCCAAAACCACACCA 23164
Db 16402 ACCGCC----------GGGCTTATGCGGAGGCTGTCGCGGGCCGTGGTGTGAGCGGG 16448
Qy 23165 CCGGACCCATCACCGGCCTGCTCTCCCTCCTCGCCCTCGACGAAACACCCCACCCCCACC 23224
Db 16449 GTCGTGTCGTTCTTGTCCTGGGATGATCGGCGGCACTCGGAGCATTCTGTTGTTCCCGCC 16508
Qy 23225 ACCCCCACACACCCACCGGCACCCTCCTCAACCTCACCCTCCCCCAAACCCACACCCAAA 23284
Db 16509 GGTCTTGCCGCGTCGCTGGTGTTGGCGCAGGCGT--TGGTTGATCTTGGCCGGGTTGGTG 16566
Qy 23285 CCCACCCACCAACCCCCCTCTGGTACGCCACCACCAACGCCACCACCACCCACCCCAACG 23344
Db 16567 AGGGGCCGCGGTTGTGGCTGGTGACGCGGGGTGCGGTGGTTGCTGGTCCTTCGGATGCCG 16626
Qy 23345 ACCCCCTCACACACCCCACCCAAGCCCAAACCTGGGGACTCGCCCGCACCACCCTCCTCG 23404
Db 16627 GTGTGGTGATTGATCCGGTGCAGGCGCAGGTGTGGGGTTTCGGGCGTGTTCTGGGTCTGG 16686
Qy 23405 AACACCCCACCCACACCGCCGGAATCATCGACCTCCCCACCACCCCCACCCCCCACACCC 23464
Db 16687 AGCATCCCGAGTTGTGGGGTGGGCTGGTGGACCTGCCGGTGGGGGTTGATGAGGAGGTGT 16746
Qy 23465 TCCACCACCTCACCCAAACCCTCACCCAACCCCACCACCAAACCCAACTCGCCATCCGCA 23524
Db 16747 GCCGGCGGTTCGTGGGTGTTGTGGCGTCGGCTGGTTTTGAGGATCAGGTGGCGGTGCGTG 16806
Qy 23525 CCACCGGCACCCACACCCGCCGCCTCACCCCCACCACCCTCACCCCCACACACCAACCAC 23584
Db 16807 GTTCGGGTGTGTGGGTGCGTCGTCTGGTGCGTGCTGTGGT------GGATGGTGGTGGGG 16860
Qy 23585 CCACCCCCACCCCCCACGGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCC 23644
Db 16861 GTGGTTGGCGGCCGCGTGGGACGGTGTTGGTCACGGGTGGTCTTGGTGGTTTGGGTGCGC 16920
Qy 23645 ACCTCACCCACCACCTCACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAA 23704
Db 16921 ATACGGCCCGGTGGTTGGTGGGTGGTGGGGCGGATCATGTGGTTCTTGTGAGCCGTCGTG 16980
Qy 23705 CCGGCCCCCACACCCCCCACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCC 23764
Db 16981 GTGGCAGTGCGCCTGGTGCTGGGGATCTGGTGCGGGAGCTGGAGGGGTTGGGCGGGGCTC 17040
Qy 23765 ACCTCACCATCACCACCTGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACA 23824
Db 17041 GGGTGTCGGTGCGGGCCTGTGATGTGGCTGATCGTGTGGCGTTGCGGGCGTTGTT----- 17095
Qy 23825 CCATCCCCCCACAACACCCCCTCACCACCGTCATCCACACCGCAGGCGTCAATCTCTTCG 23884
Db 17096 -GTCGGATCTGGGTGAGCCGGTGACGGCGGTGTTCCATGCGGCTGGTGTTCCTCAGTCGA 17154
Qy 23885 CCCCCGTGTCGGAAACCGATGCCGAATCCTTCTCTTCCGTTACGGCAGCGAAGGCAACGG 23944
Db 17155 CGCCTTTGGCGGAGATCTCTGTCCAGGAGGCGGCTGATGTGATGGCGGCCAAGGTGGCGG 17214
Qy 23945 GCGCGGCGATTCTGCATGAGTTGCTGCTGGACCATGAAACGCTTGAACACTTCATTCTCT 24004
Db 17215 GTGCGGTGAATCTGGGTGAGTTGGTGGATCCCTGTG---GTCTGGAGGCGTTTGTGTTGT 17271
Qy 24005 TCTCGTCGGGCGCCGGCGCTTGGGGCAGCGGGAATCAGTGCGCATACTCGGCGGCCAACG 24064
Db 17272 TCTCCTCCAATGCCGGTGTGTGGGGCAGTGGGGGGCAGGCGGTGTATGCGGCGGCGAATG 17331
Qy 24065 CATACCTGGACGCGCTCGCGACGCATCGTCAGACACATGGACTTCCCGGGGCATCGATCG 24124
Db 17332 CGTTTCTTGATGCGTTGGCGGTGCGTCGTCGGGGTGTTGGTCTGCCGGCCACGAGTGTGG 17391
Qy 24125 CCTGGGGCCCCTGGGCCGGAAAGGGCATGTCGGCCGGTGATGCGGCTCATGGTTACCTGG 24184
Db 17392 CGTGGGGGATGTGGGCTGGTGAGGGGATGGCGTCGGTGGGTGGTGCGGCGCGGGAGTTGT 17451
Qy 24185 AAAAGCGCGGCATTCTGCCGATGGAGCCACGCATGGCGCTCGCGGCATTCCATCGTGCGC 24244











```
Qy      11234 CACTGGCCCTCTTCGATGCGGCCCTGGCTCTGGACCGGCCGGTCCTGCTGCCCGCCGACC 11293
Db      12929 GCATCGCGCTCCTCGACGCCGCCCTCCGCGACGACCGCCACCCGGTCCTGCTGCCCCTGC 12988

Qy      11294 TGCGTCCCGCCCCGCCCCTGCCG---------------------------------- 11316
Db      12989 GGCTCGACGCCGCCGGGCTGCGGGACGCGGCCGGGAACGACCCGGCCGGAATCCCGGCGC 13048

Qy      11317 ------------CCCCTGCTGCAGGACCTCCTGCCCGCCACCCGCCGCCGCACCACCCG 11363
Db      13049 TCTTCCGGGACGTCGTCGGCGCCAGGACCGTCCGGGCCCGGCCGTCCGCGGCCTCCGCCT 13108

Qy      11364 CACCACCACTACCGGTGGTGCGGACAACGGCGCCCA---------------------- 11399
Db      13109 CGACGACAGCCGGGACGGCCGGCACGCCGGGGACGGCGGACGGCGCGGCGGAAACGGCGG 13168

Qy      11400 -------GCTGCACGCCCGGCTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCC 11452
Db      13169 CGGTCACGCTCGCCGACCGGGCCGCCACCGTGGACGGGCCCGCACGGCAGCGCCTGCTGC 13228

Qy      11453 TCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCC 11512
Db      13229 TCGAGTTCGTCGTCGGCGAGGTCGCCGAAGTACTCGGCCACGCCCGCGGTCACCGGATCG 13288

Qy      11513 CCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCA 11572
Db      13289 ACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGACCGCCGTCGAACTCCGCA 13348

Qy      11573 ACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCA 11632
Db      13349 ACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGCGACCCTGGTCTTCGACCACCCAA 13408

Qy      11633 ACCCCACCACCCTCACCCACCACCTCCACACACAACTCCAGC 11674
Db      13409 GCCCGGCGGCACTCGCCTCCCACCTGGACGCCGAGCTGCCGC 13450
```

RESULT 10
US-09-808-880-1
; Sequence 1, Application US/09808880
; Publication No. US20030027287A1
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/808,880
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US/09/428,517
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/120,254
; PRIOR FILING DATE: 1999-02-16
; PRIOR APPLICATION NUMBER: 60/106,100
; PRIOR FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 50937
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artifical Sequence: Recombinant DNA
US-09-808-880-1

Query Match 7.5%; Score 2304.2; DB 9; Length 50937;
Best Local Similarity 50.8%; Pred. No. 0;
Matches 7910; Conservative 0; Mismatches 6603; Indels 1044; Gaps 60;

```
Qy      12260 GCGGAGGGTTTCTCTATGACGCGGGCGACTTCGACGCCGCGTTCTTCGAGTTGTCACCGC 12319
Db       5953 GCGGCGGATTCCTCGACAGCATCGACACTTTCGACGCGGATTTCTTCAACATCTCGCCCA 6012

Qy      12320 GTGAGGCGCTGGCAATGGACCCGCAGCAGCGCCTGCTGCTCGAAACCACTTGGGAAACGT 12379
Db       6013 GAGAAGCCGGTGTCCTCGACCCCCAGCAACGCCTCGCGCTGGAACTCGGCTGGGAGGCGC 6072

Qy      12380 TCGAACAGGGCGGAATCGACCCGAGGTCCATGCGCGGAAGCCGGACCGGGGTTTTCGTGG 12439
Db       6073 TGGAAGACGCCGGAATCGTCCCGCGACACCTCAGGGGAACCCGCACCTCGGTCTTCATGG 6132

Qy      12440 GGATCAATCCGGAGGACTACACCACCGGATACACACATCAGCCCTCAAACGCAGTCGAGG 12499
Db       6133 GCGCCATGTGGGACGACTACGCGCACCTGGCGCACGCACGGGGAGAAGCCGCCCTCACCC 6192

Qy      12500 GCTACCTGCTCACTGGCAGCGCGGCAAGCATTGCGTCAGGCCGTATCTCCTACAACTTCG 12559
Db       6193 GGCATTCCCTGACGGGAACGCACCGCGGCATGATCGCCAACCGGCTCTCCTACGCCCTGG 6252

Qy      12560 GGCTCGAAGGCCCTGCGATCACTATCGACACCGCGTGTTCCTCCTCGCTCGTCGCCCTGC 12619
Db       6253 GCCTCCAAGGCCCCAGCCTCACCGTCGACACCGGACAATCCTCCTCCCTCGCCGCCGTGC 6312

Qy      12620 ATCTGGCCTGCCAAGCGCTCCGGTCCGGTGAATGCACCATGGCGCTCGCAGGCGGCGCCT 12679
Db       6313 ACATGGCCTGCGAGAGCCTGGCCCGCGGCGAATCCGACCTCGCCCTCGTCGGCGGCGTCA 6372

Qy      12680 CCGTCATGGCCACTCCCTTCGTCTTCACCGAGTTCTCTCGCCAGCGGGGCCTGGCCGCAG 12739
Db       6373 ACCTCGTCCTCGATCCGGCCGGCACGACCGGCGTCGAGAGGTTCGGAGCACTCTCACCGG 6432

Qy      12740 ACGGCCGGTGCAAGGCGTTTTCGGCGGCGGCGGACGGGACCGGCTGGTCCGAGGGTGTGG 12799
Db       6433 ACGGCAGGTGCTACACCTTCGACTCCCGGGCGAACGGCTACGCCCGAGGAGAGGGCGGCG 6492

Qy      12800 GGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCGCCGCAACGGTCACCGTGTCCTGGCCG 12859
Db       6493 TCGTAGTCGTCCTCAAGCCCACCCACCGCGCGCTCGCGGACGGTGACACCGTCTACTGCG 6552

Qy      12860 TCGTCCGCGGCAGCGCCGTCAACCAGGACGGCGCAAGCAACGGCCTGACCGCACCCAACG 12919
Db       6553 AGATCCTGGGCAGCGCCCTCAACAACGACGGCGCCACGGAAGGCCTCACCGTCCCCAGCG 6612

Qy      12920 GTCGTTCACAAGTCAAGGTCATCCGCCAGGCTTTGGCCAACGCACACCTCTCCCCTGCCG 12979
Db       6613 CCCGCGCCCAGGCGGACGTCCTGCGACAGGCATGGGAACGGGCACGCGTGGCCCCGACGG 6672

Qy      12980 ATGTCGATGCGGTGGAGGCCCACGGCACGGGGACCACCCTGGGCGACCCGATCGAGGCTC 13039
Db       6673 ACGTCCAGTACGTGGAACTGCACGGAACCGGCACACCGGCCGGCGACCCCGTCGAGGCCG 6732

Qy      13040 AAGCCCTCGTCGAAGCCTACGGTCAGGACCGCCCCAACGGCCGCCCCCTCTGGCTCGGAA 13099
Db       6733 AGGGCCTCGGCACCGCGCTCGGCACCGCACGCCCGGCCGAGGCGCCGCTCCTGGTCGGCT 6792

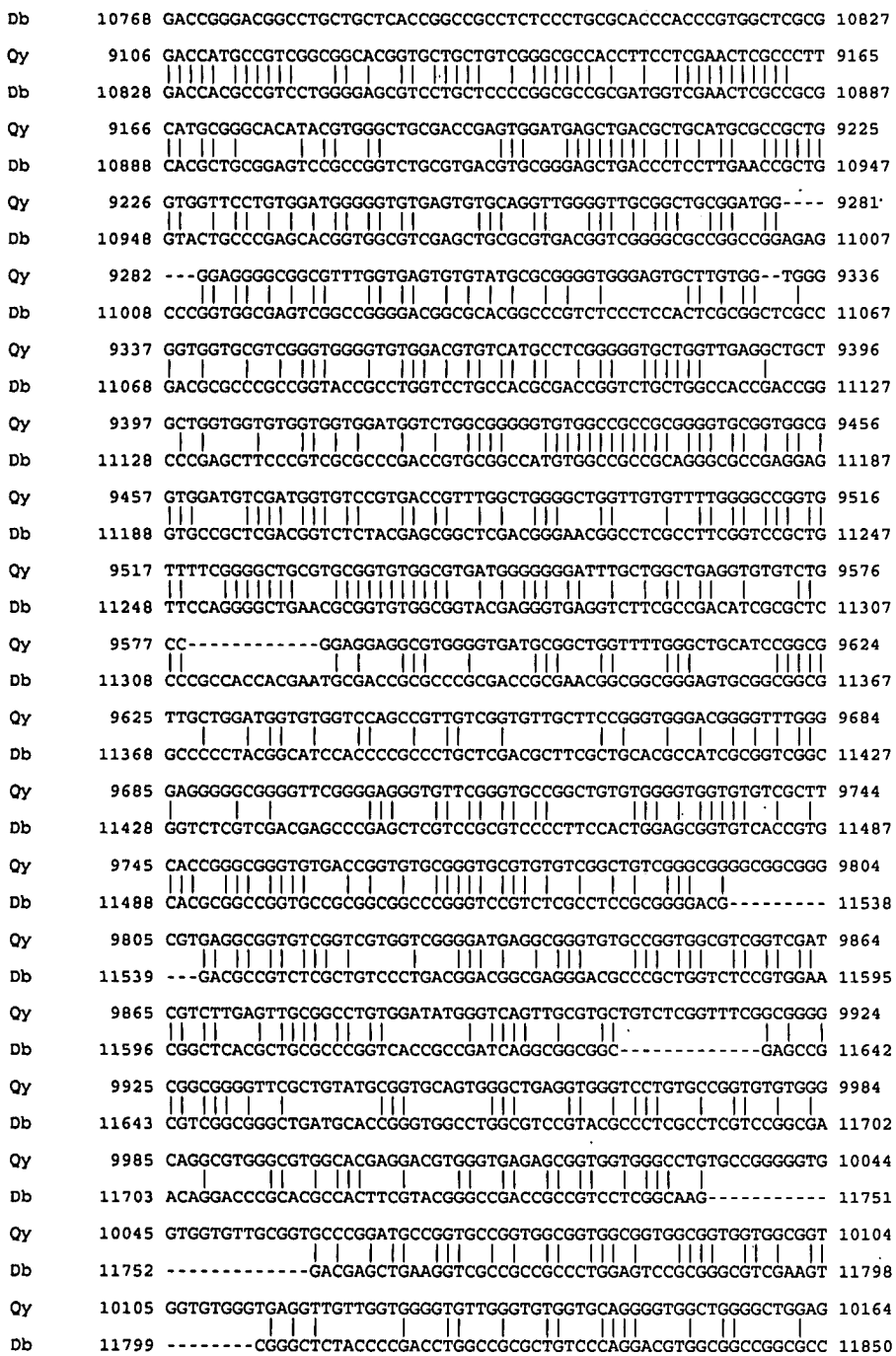
Qy      13100 CCCTCAAGTCCAACATCGGGCACTCCATGGCCGCTGCGGGTGTGGGCGGGGTCATCAAGA 13159
Db       6793 CGGTCAAGACGAACATCGGTCACCTCGAAGGCGCGGCAGGCATCGCGGGCCTCCTGAAGA 6852

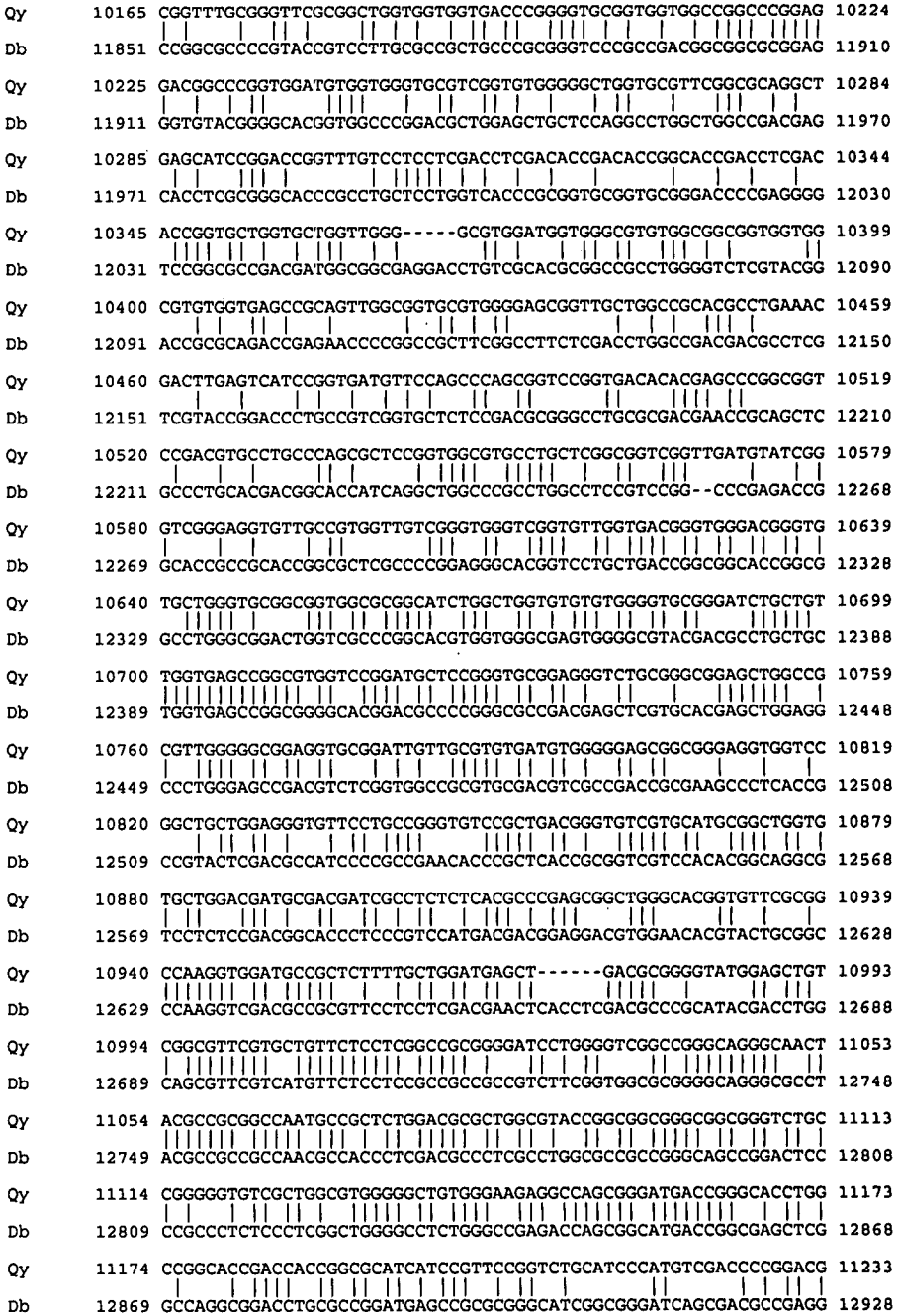
Qy      13160 TGGTGATGGCGCTGCGGAATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCGC 13219
Db       6853 CGGTCCTGAGCATCAAGAACCGGCACCTCCCGGCAAGCCTGAACTTCACCTCGCCCAACC 6912

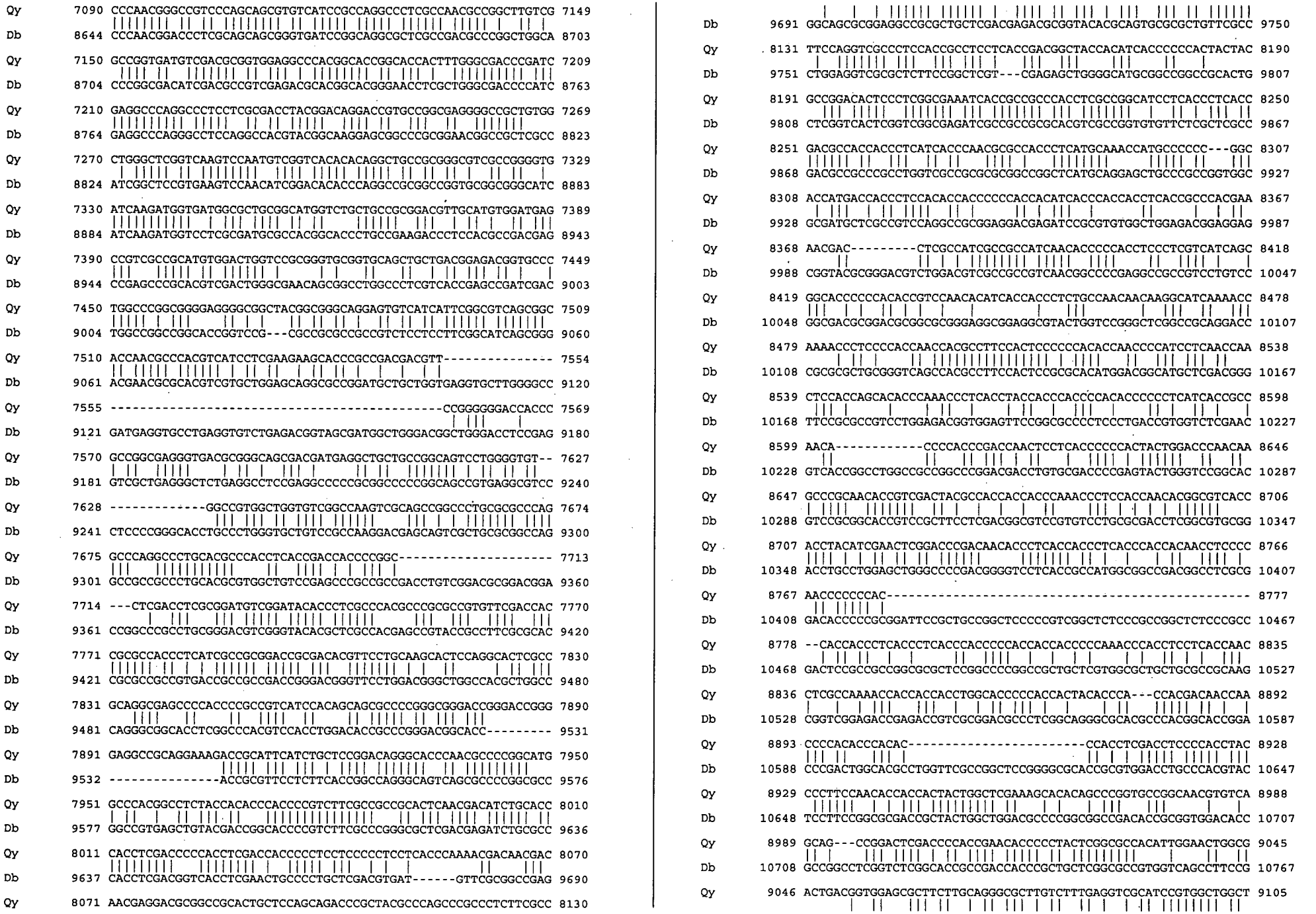
Qy      13220 CGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCCG 13279
Db       6913 CCCGCATCGACCTCGACGCCCTGCGCCTGCGCGTCCACACCGCGTACGGCCCCTGGCCGA 6972

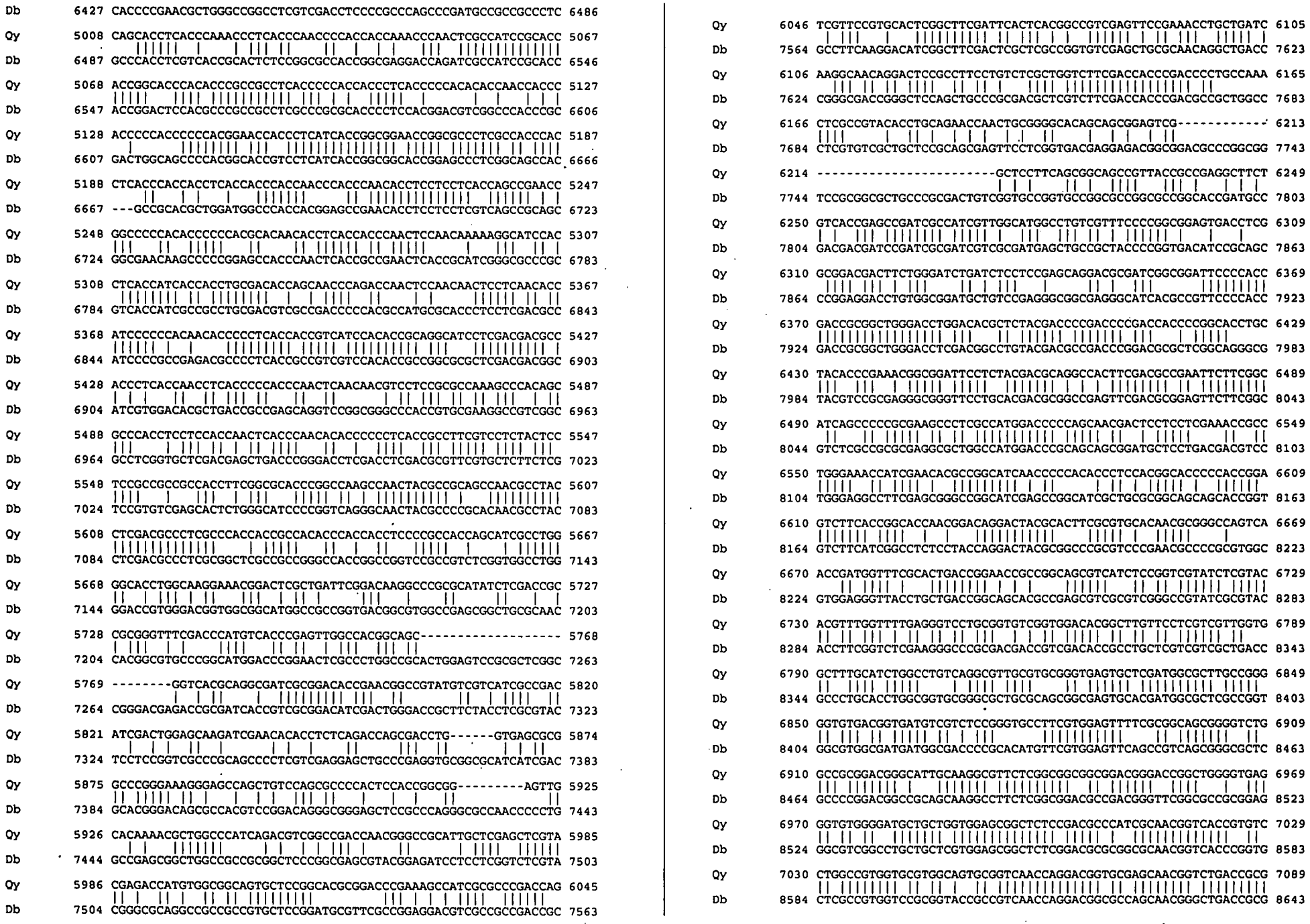
Qy      13280 GCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAACG 13339
Db       6973 GCCCCGACCGGCCGCT---GGTGGCGGGCGTCTCCTCCTTCGGCATGGGCGGGACGAACT 7029

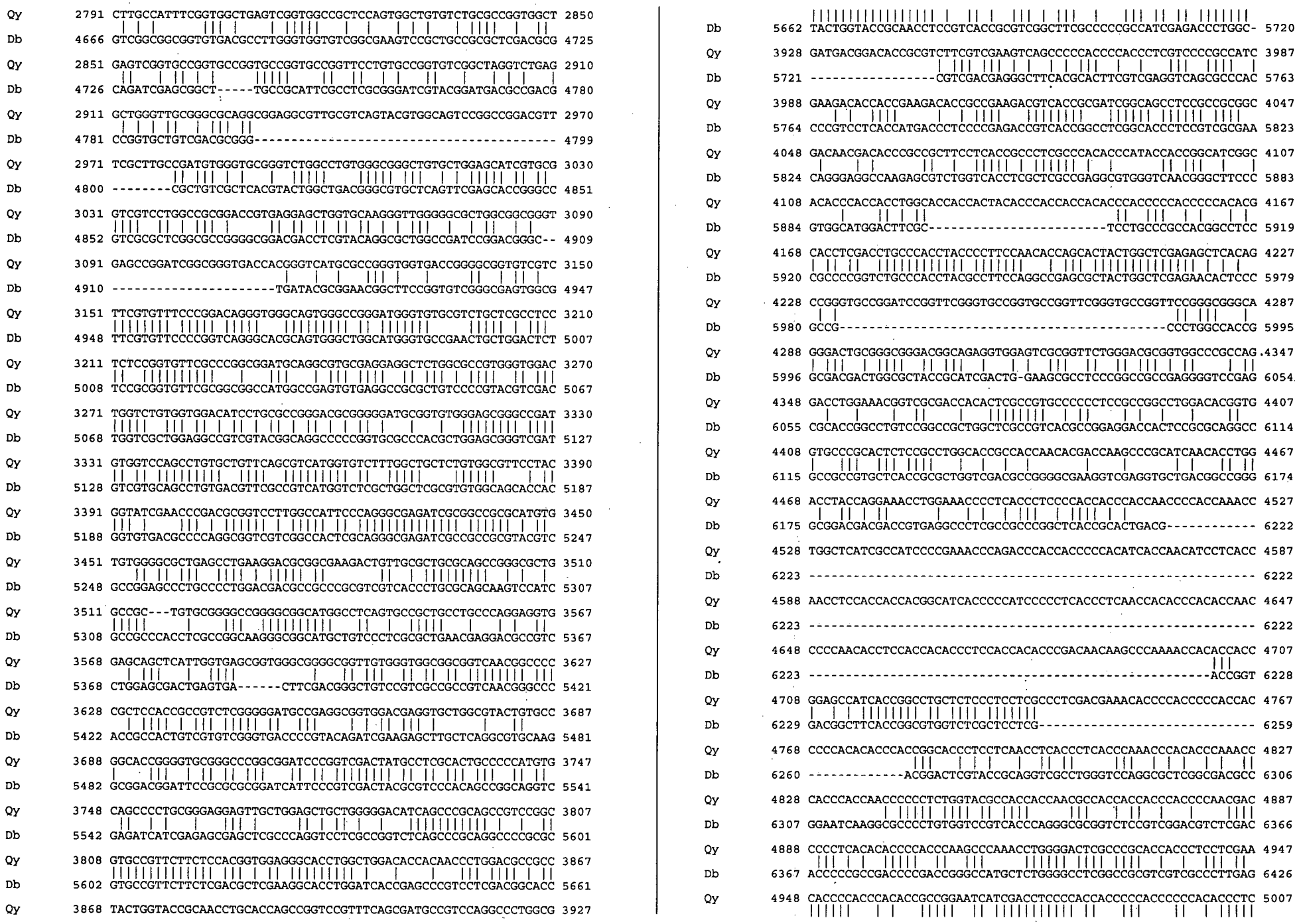
Qy      13340 CCCACGTCATCCT--------------CGAGGAAGCACCCGCCCACAACATCCCGTCAG 13384
Db       7030 GCCACGTCGTCCTGTCCGAGTTACGGAACGCGGGAGGCGACGGCGCCGGAAAAGGGCCGT 7089
```

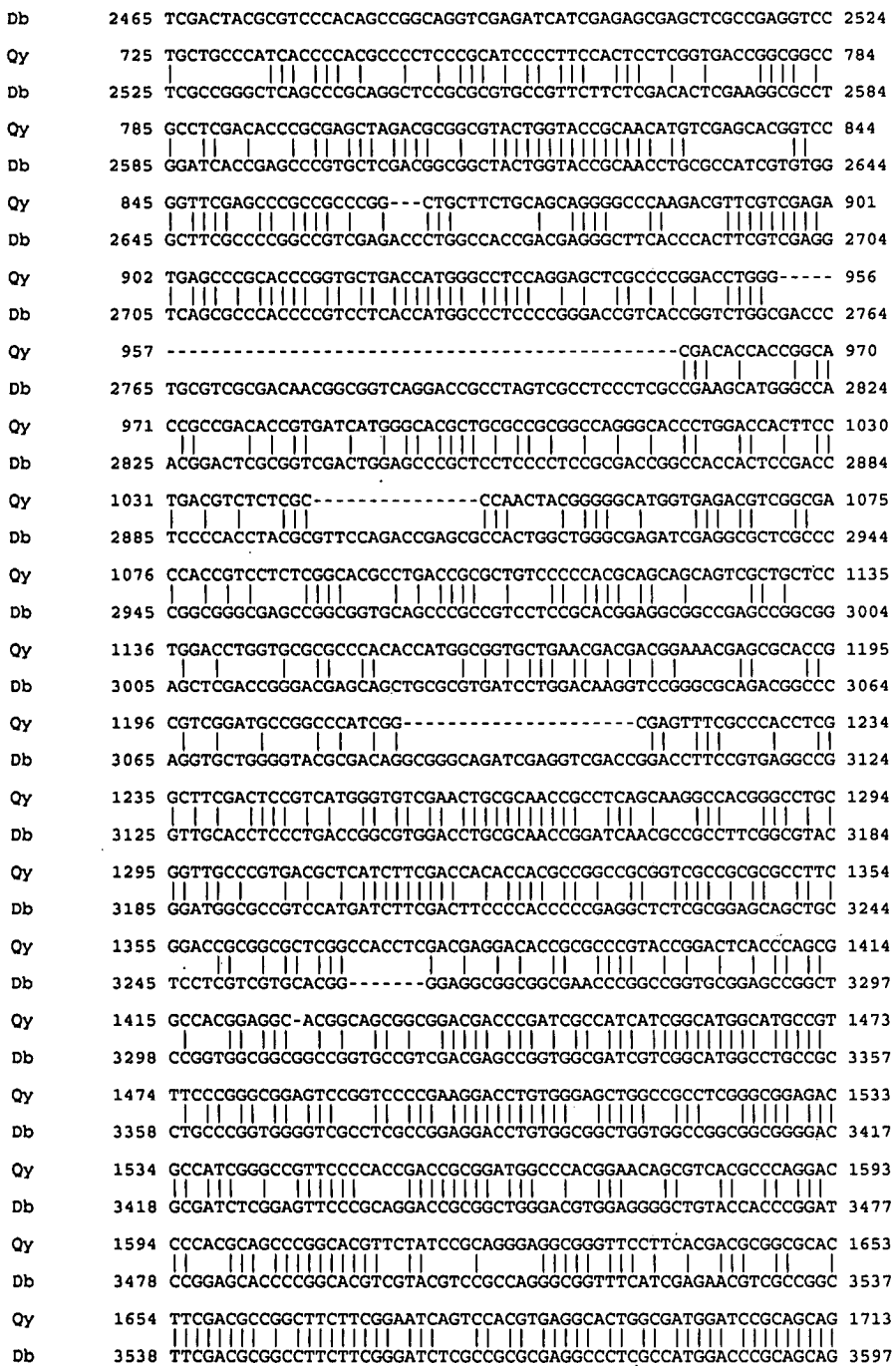


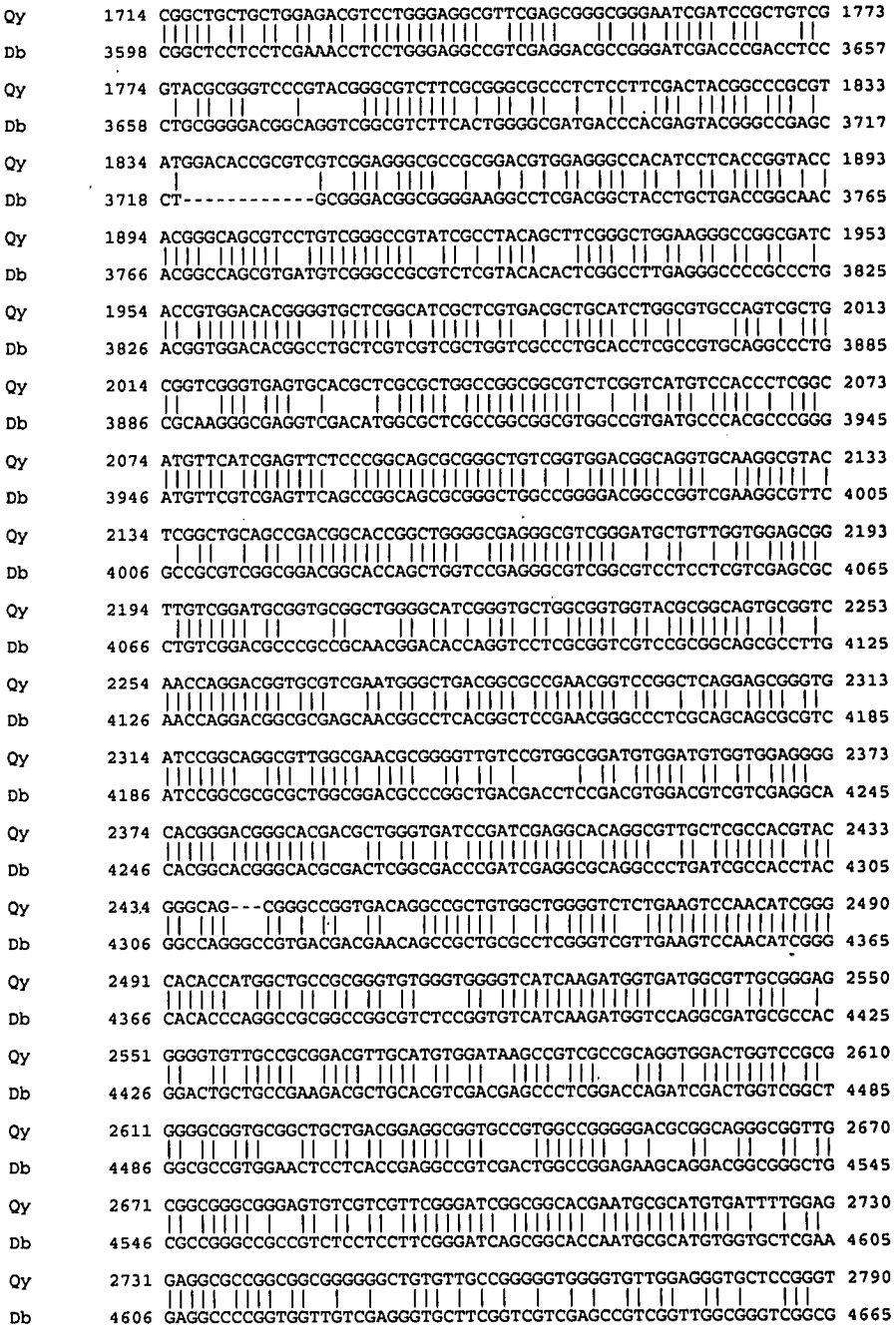












```
Qy      10700 TGGTGAGCCGGCGTGGTCCGGATGCTCCGGGTGCGGAGGGTCTGCGGGCGGAGCTGGCCG 10759
              ||||||||||||| ||  |||| || ||||| || || |  ||   |   |||||||  |
Db      12389 TGGTGAGCCGGCGGGGCACGGACGCCCCGGGCGCCGACGAGCTCGTGCACGAGCTGGAGG 12448

Qy      10760 CGTTGGGGGCGGAGGTGCGGATTGTTGCGTGTGATGTGGGGGAGCGGCGGGAGGTGGTCC 10819
              |  |||| || || ||    | | |  ||||| || || |  || ||     |   |   |
Db      12449 CCCTGGGAGCCGACGTCTCGGTGGCCGCGTGCGACGTCGCCGACCGCGAAGCCCTCACCG 12508

Qy      10820 GGCTGCTGGAGGGTGTTCCTGCCGGGTGTCCGCTGACGGGTGTCGTGCATGCGGCTGGTG 10879
                | || || |   | || ||||     ||||| || |  ||||| ||  |||| || |
Db      12509 CCGTACTCGACGCCATCCCCGCCGAACACCCGCTCACCGCGGTCGTCCACACGGCAGGCG 12568

Qy      10880 TGCTGGACGATGCGACGATCGCCTCTCTCACGCCCGAGCGGCTGGGCACGGTGTTCGCGG 10939
              | ||   ||| |  ||  || | ||  | ||| | |||    |||     ||  |   |
Db      12569 TCCTCTCCGACGGCACCCTCCCGTCCATGACGACGGAGGACGTGGAACACGTACTGCGGC 12628

Qy      10940 CCAAGGTGGATGCCGCTCTTTTGCTGGATGAGCT------GACGCGGGGTATGGAGCTGT 10993
              ||||||| || |||||  |  | || || || ||      |||||  |      || |||
Db      12629 CCAAGGTCGACGCCGCGTTCCTCCTCGACGAACTCACCTCGACGCCCGCATACGACCTGG 12688

Qy      10994 CGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATCCTGGGGTCGGCCGGGCAGGGCAACT 11053
              | ||||||||  |||||||||| ||||| |   || | ||    || |||||||||  ||
Db      12689 CAGCGTTCGTCATGTTCTCCTCCGCCGCCGCCGTCTTCGGTGGCGCGGGGCAGGGCGCCT 12748

Qy      11054 ACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCGTACCGGCGGCGGGCGGCGGGTCTGC 11113
              ||||||| ||||| ||| | || ||||| || || |  || || ||||| || || || |
Db      12749 ACGCCGCCGCCAACGCCACCCTCGACGCCCTCGCCTGGCGCCGCCGGGCAGCCGGACTCC 12808

Qy      11114 CGGGGGTGTCGCTGGCGTGGGGGCTGTGGGAAGAGGCCAGCGGGATGACCGGGCACCTGG 11173
              |  |   | || || |  ||||| || ||||  ||| ||||||| ||||||||  | || |
Db      12809 CCGCCCTCTCCCTCGGCTGGGGCCTCTGGGCCGAGACCAGCGGCATGACCGGCGAGCTCG 12868

Qy      11174 CCGGCACCGACCACCGGCGCATCATCCGTTCCGGTCTGCATCCCATGTCGACCCCGGACG 11233
              |  |  ||||  || || || | |||  | ||  |      ||      | | || | |
Db      12869 GCCAGGCGGACCTGCGCCGGATGAGCCGCGCGGGCATCGGCGGGATCAGCGACGCCGAGG 12928

Qy      11234 CACTGGCCCTCTTCGATGCGGCCCTGGCTCTGGACCGGCCGGTCCTGCTGCCCGCCGACC 11293
                | || ||| |||| || |||||       |||||  |     | ||||   ||   |
Db      12929 GCATCGCGCTCCTCGACGCCGCCCTCCGCGACGACCGCCACCCGGTCCTGCTGCCCCTGC 12988

Qy      11294 TGCGTCCCGCCCCGCCCCTGCCG---------------------------------------- 11316
               ||    |||| |    |||| |
Db      12989 GGCTCGACGCCGCCGGGCTGCGGGACGCGGCCGGGAACGACCCGGCCGGAATCCCGGCGC 13048

Qy      11317 ------------CCCCTGCTGCAGGACCTCCTGCCCGCCACCCGCCGCCGCACCACCCG 11363
                         | | ||  |||||||  | |  | |  ||  |||| ||  |  ||
Db      13049 TCTTCCGGGACGTCGTCGGCGCCAGGACCGTCCGGGCCCGGCCGTCCGCGGCCTCCGCCT 13108

Qy      11364 CACCACCACTACCGGTGGTGCGGACAACGGCGCCCA---------------------- 11399
              |  |  ||      |  ||   | ||  ||| | |
Db      13109 CGACGACAGCCGGGACGGCCGGCACGCCGGGGACGGCGGACGGCGCGGCGGAAACGGCGG 13168

Qy      11400 -------GCTGCACGCCCGGCTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCC 11452
                    |||   || |||   |||  |  |              || |||   |   || |
Db      13169 CGGTCACGCTCGCCGACCGGGCCGCCACCGTGGACGGGCCCGCACGGCAGCGCCTGCTGC 13228

Qy      11453 TCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCC 11512
              |||   | |||  |  | |  |||||   || || ||||||  ||  |    ||   |||
Db      13229 TCGAGTTCGTCGTCGGCGAGGTCGCCGAAGTACTCGGCCACGCCCGCGGTCACCGGATCG 13288

Qy      11513 CCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCA 11572
              | |||| || |  |||| |||||||||||||||||||||||||||||||||||||||||
Db      13289 ACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGACCGCCGTCGAACTCCGCA 13348

Qy      11573 ACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCA 11632
              |||||||  |  | ||   || |||  ||||||  | |||||  | ||||||||||||  |
Db      13349 ACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGCGACCCTGGTCTTCGACCACCCAA 13408

Qy      11633 ACCCCACCACCCTCACCCACCACCTCCACACACAACTCCAGC 11674
               |||  |  | ||| ||  ||||||  || |  | || | ||
Db      13409 GCCCGGCGGCACTCGCCTCCCACCTGGACGCCGAGCTGCCGC 13450
```

RESULT 9
US-09-861-289-30
; Sequence 30, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-30

Query Match 7.6%; Score 2342; DB 10; Length 13842;
Best Local Similarity 54.3%; Pred. No. 0;
Matches 6606; Conservative 0; Mismatches 4530; Indels 1026; Gaps 52;

```
Qy          8 GGATGGACGGCGGGGAAGAACCCCGCCCTGCGGCAGGGGAGGTCCTCGGAGTGGCCGACG 67
              ||  |||||     ||  | | | |    | ||| | | ||    |  || | |  |  |
Db       1820 GGCCGGACGATCTGGCGGCAGCGCTGGCCGCGCCTGAGGGTCTGGTCCGGGGCGTGGCTT 1879

Qy         68 AGGCGGACGGCGGCGTCGTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGG 127
                |  | |||  | || |  ||||| || |||||||||||| |||| ||| | |||||||
Db       1880 CCGGTGTCGGGCGAGTGGCGTTCGTGTTCCCCGGGCAGGGCACGCAGTGGGCCGGCATGG 1939

Qy        128 GAAGGGAACTTCTCGACGCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCG 187
              |    ||||| || ||| ||||||  || |||  || |  | |   ||  ||||| ||||
Db       1940 GTGCCGAACTGCTGGACTCTTCCGCGGTGTTCGCGGCGGCCATGGCCGAATGCGAGGCCG 1999

Qy        188 CGTTCGCGCCCTACGTCGACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTC 247
              |  || | ||  |||||||||||||||| |||||    ||   | ||| |  | || | || |
Db       2000 CACTCTCCCCGTACGTCGACTGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGC 2059

Qy        248 CCGGGCTGGACCGGGTGGACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGG 307
              ||  |||||| ||||| || ||||| |||||       |||||||||||| |||| ||||
Db       2060 CCACGCTGGAGCGGGTCGATGTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGG 2119

Qy        308 CCGCCCTCTGGCGCTCGCAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCG 367
              |   | | ||||     || |||||   |||    |||||  | || |||   | |||||
Db       2120 CTCGCGTGTGGCAGCACCACGGGGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCG 2179

Qy        368 AGATCGCGGCAGCCCACGTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGA 427
              ||||||| || ||  ||||| | || | ||||  |||||  |||||||||  || || ||||
Db       2180 AGATCGCCGCCGCGTACGTCGCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGA 2239

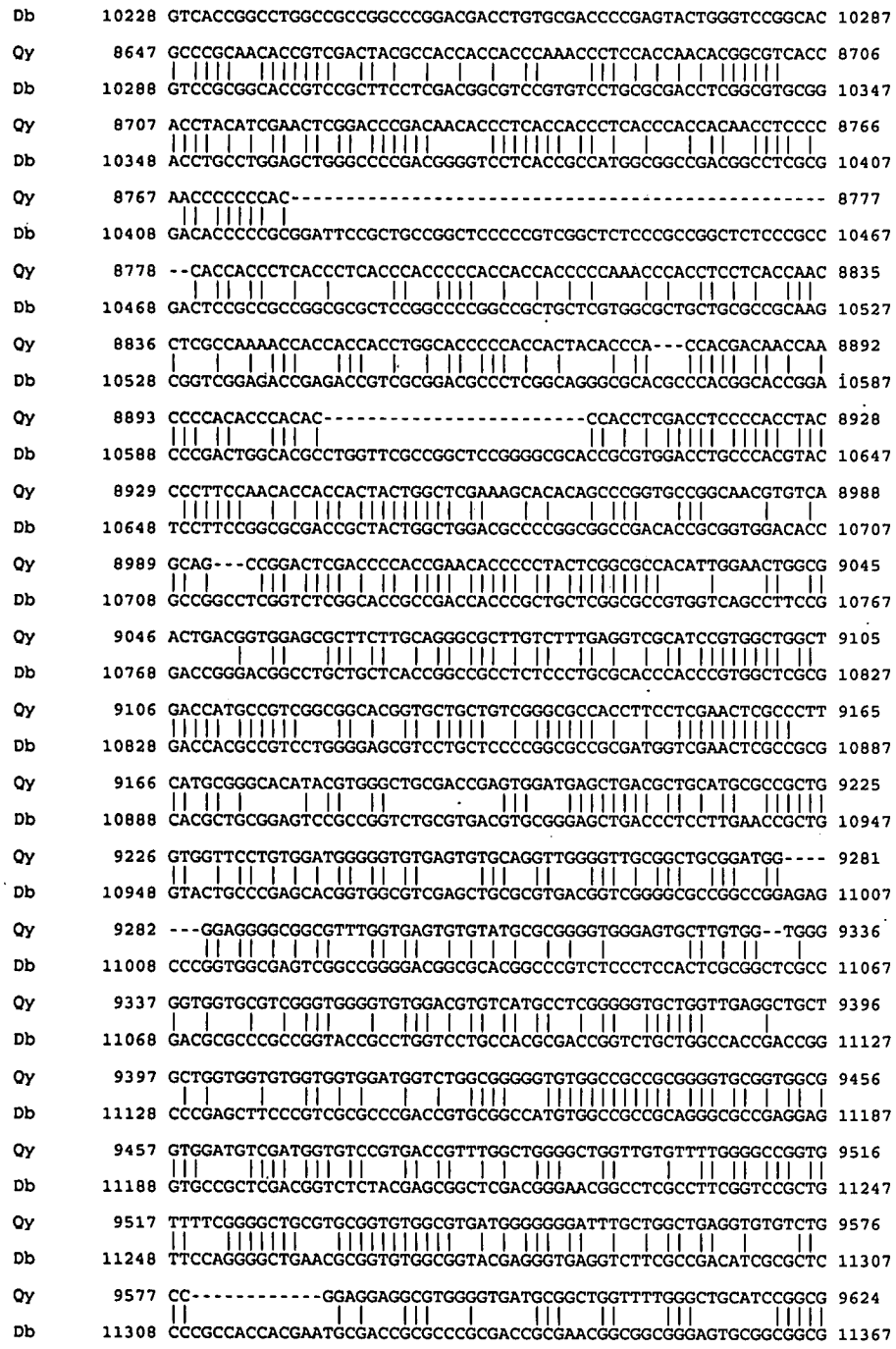
Qy        428 CGCTTTGGAGCCAGGCACAGACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCG 484
              | ||  | ||| || |      || |   ||| ||||| |  ||||   |  | ||| |||
Db       2240 CCCTGCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCG 2299

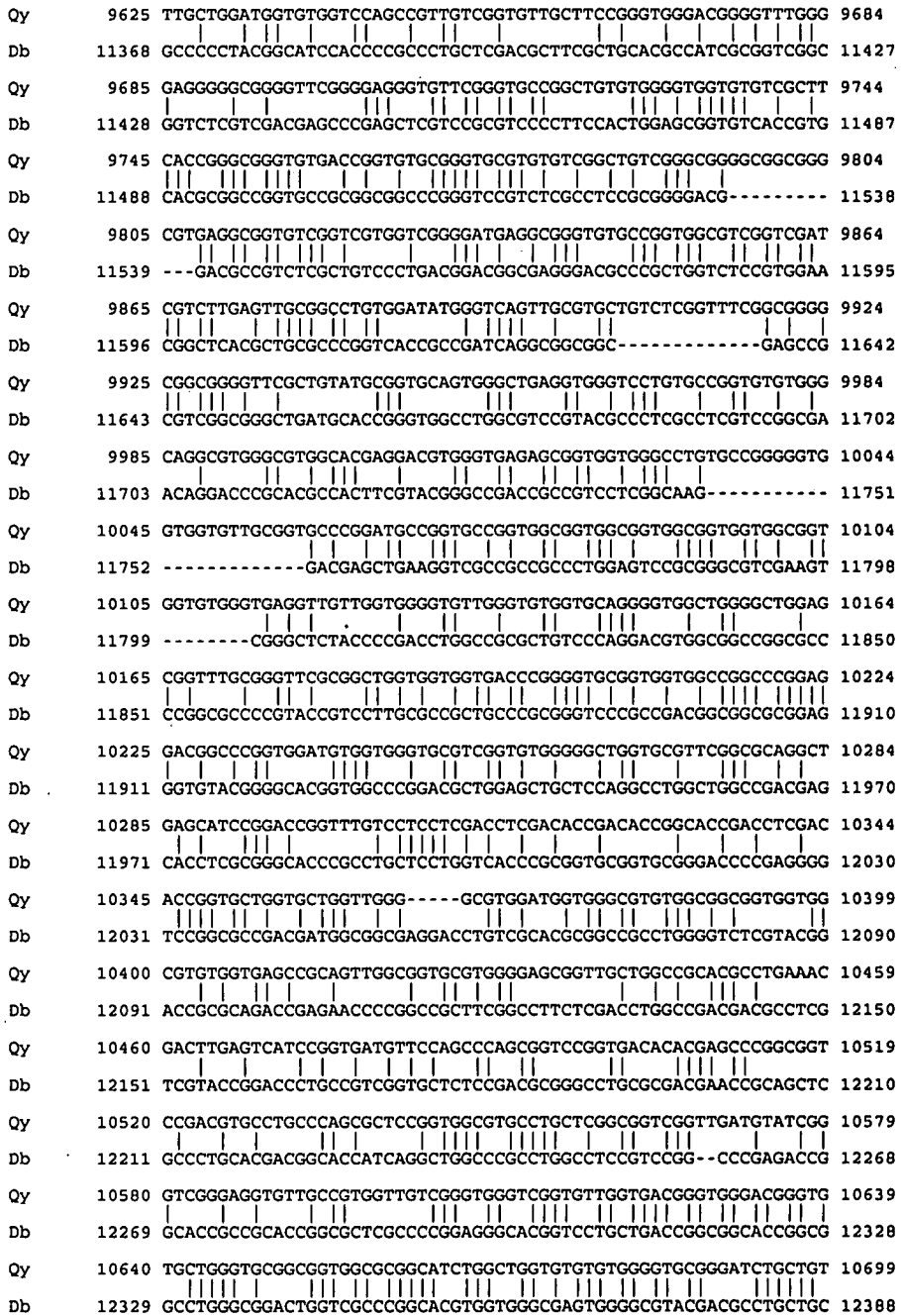
Qy        485 CCGCCACGCCGGATGAGCTCCTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGC 544
              |    |   ||| |  |||||  ||| | ||   ||   ||
Db       2300 CGCTGAGCGAGGACGCCGTCCTGGAGCGACTGGCCGGGTTCGACG--------------- 2344

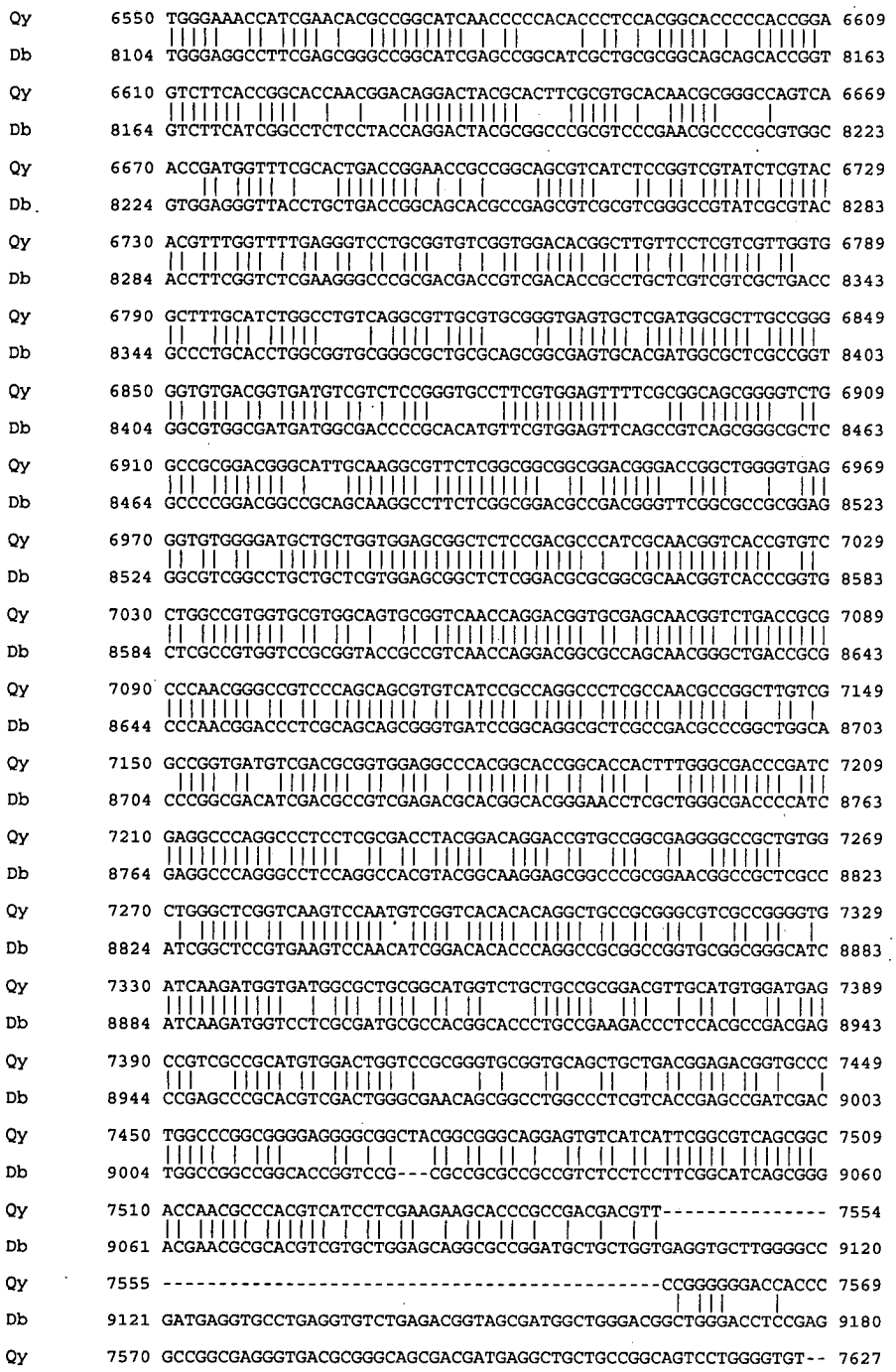
Qy        545 GGCTCGCCGTCGCAGCCGTCAACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGG 604
              ||||  |||||| || || ||||| |||      ||| || || |||||| || ||
Db       2345 GGCTGTCCGTCGCCGCTGTGAACGGGCCCACCGCCACCGTGGTCTCCGGTGACCCCGTAC 2404

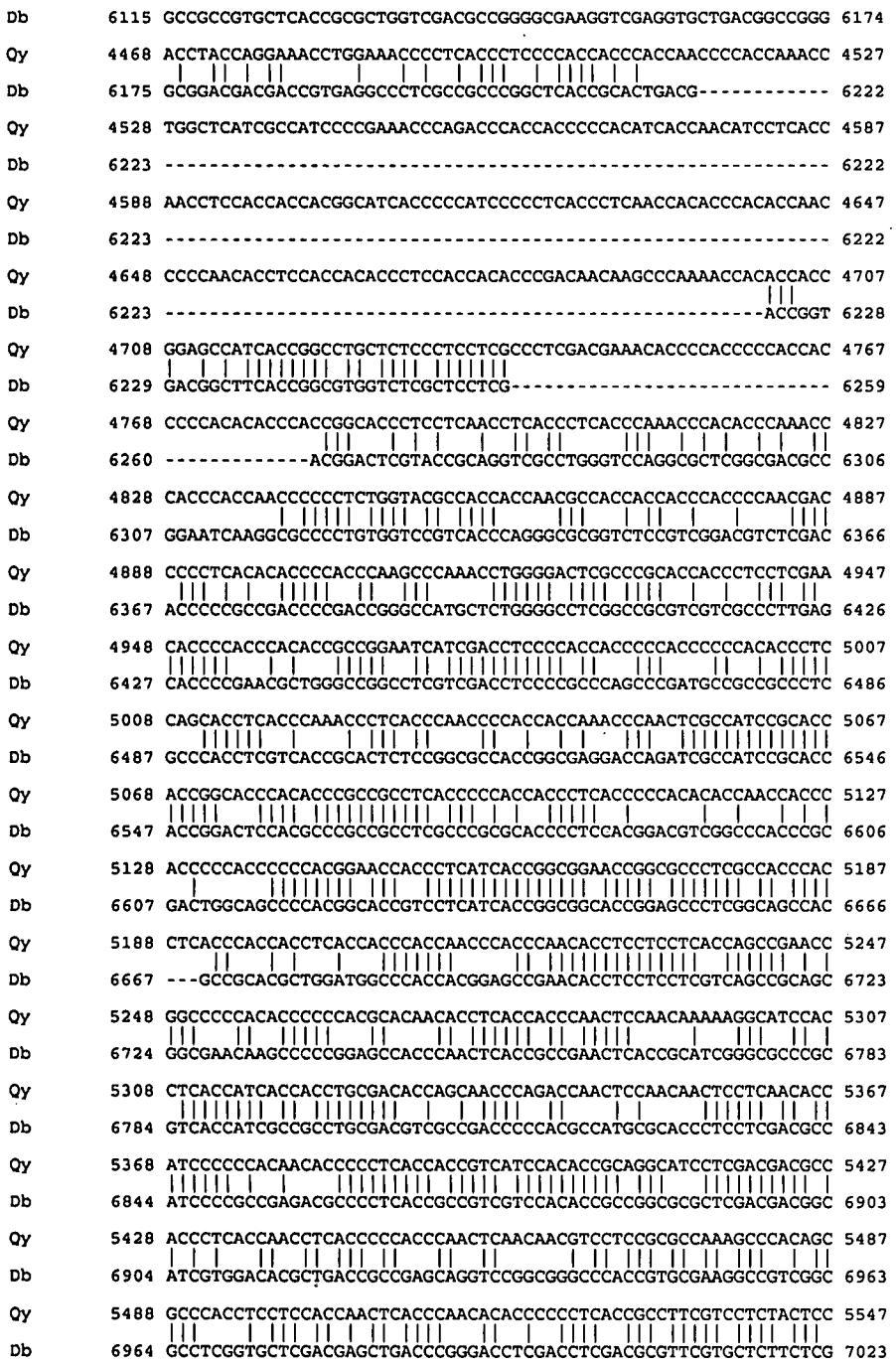
Qy        605 CCGTCGCGGACCTGGTGGCCGACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGG 664
              |||  || || |    |         ||||    ||| ||  ||||  | || || |
Db       2405 AGATCGAAGAGCTTGCTCGGGCGTGTGAGGCCGATGGGGTCCGTGCGCGGGTCATTCCCG 2464

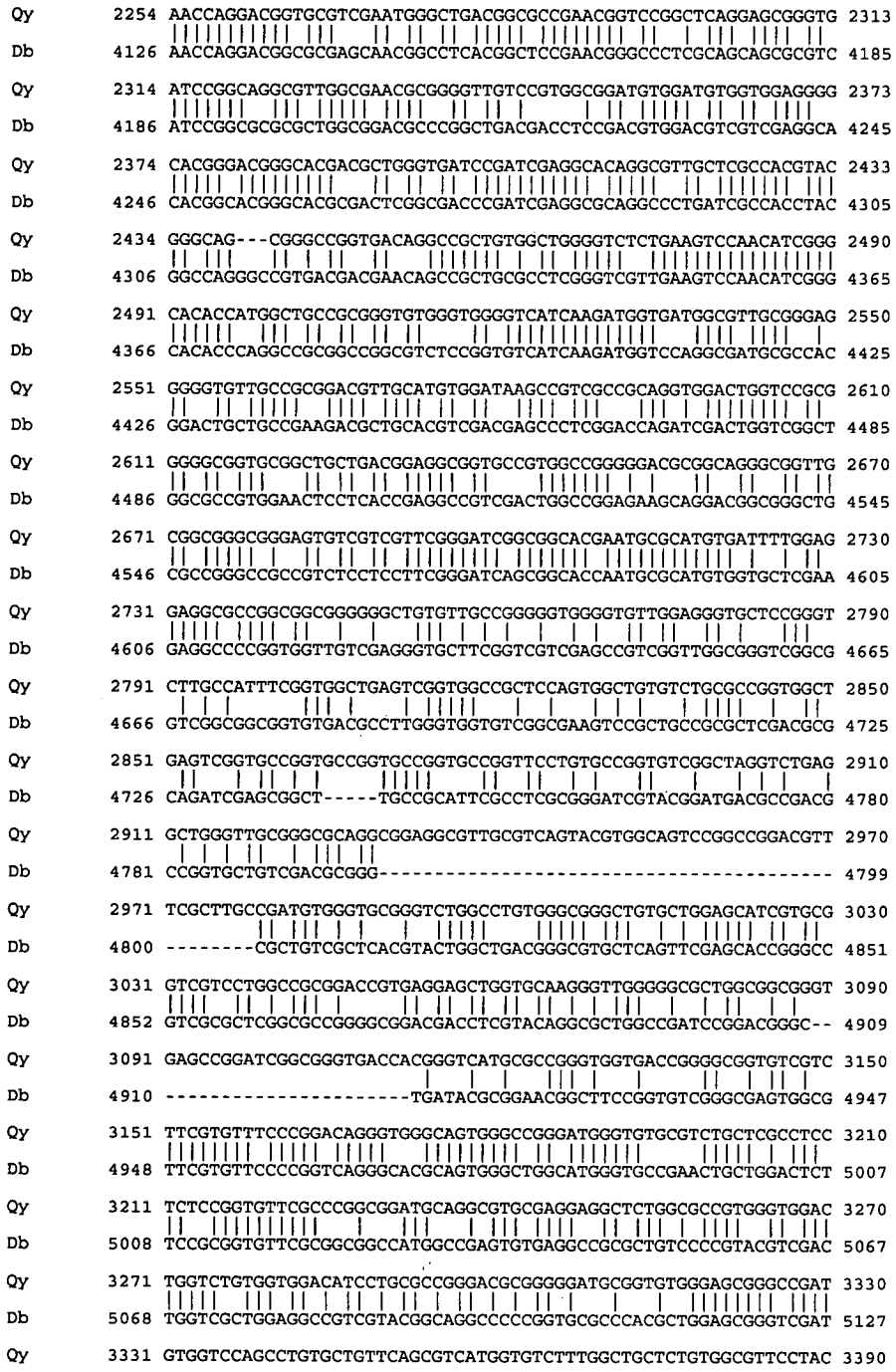
Qy        665 TGGACGTTCCCGCCCACTCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCC 724
              | |||    |  |||||  ||  | | |  |   |||||||     | ||| |   | ||
```

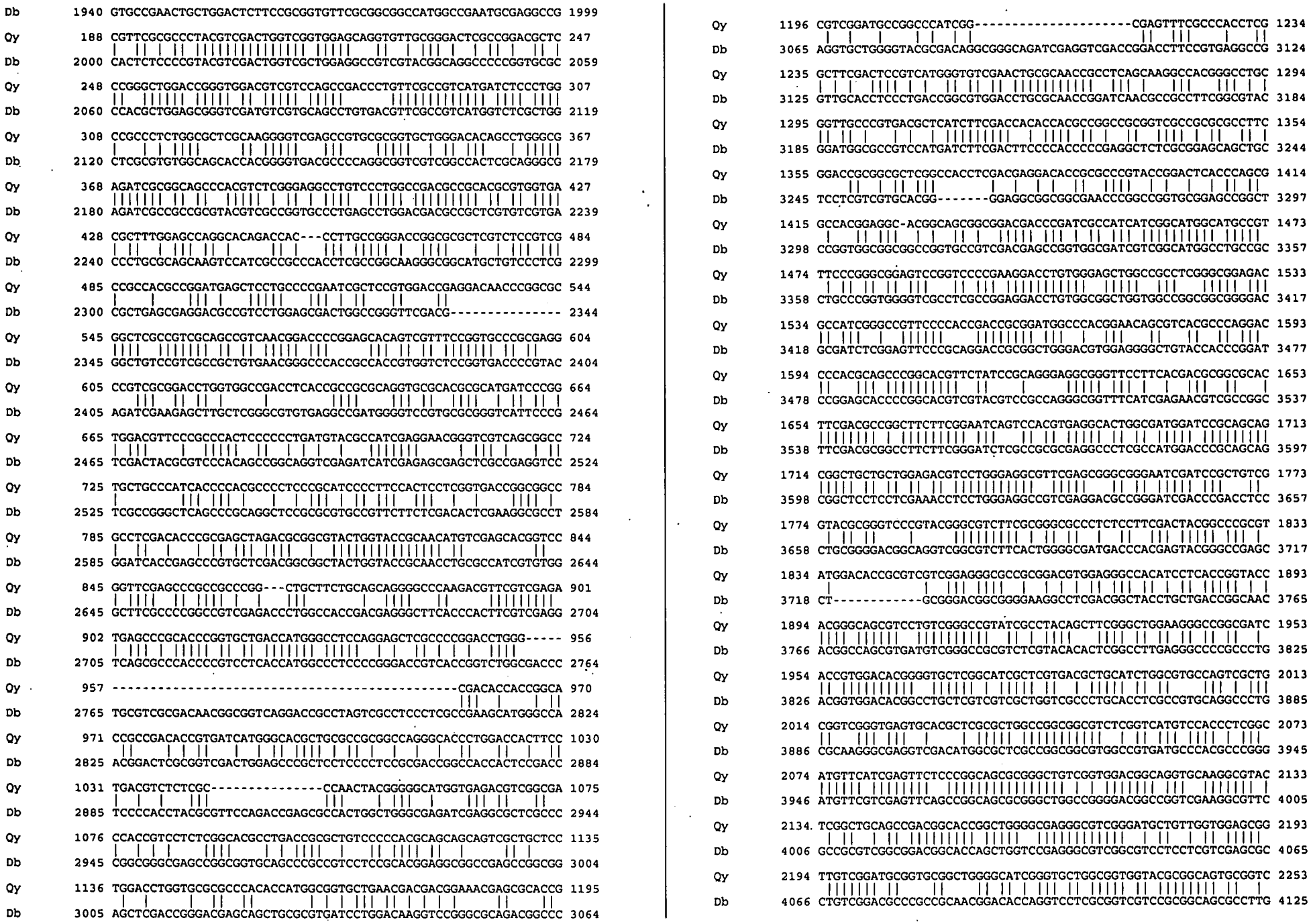


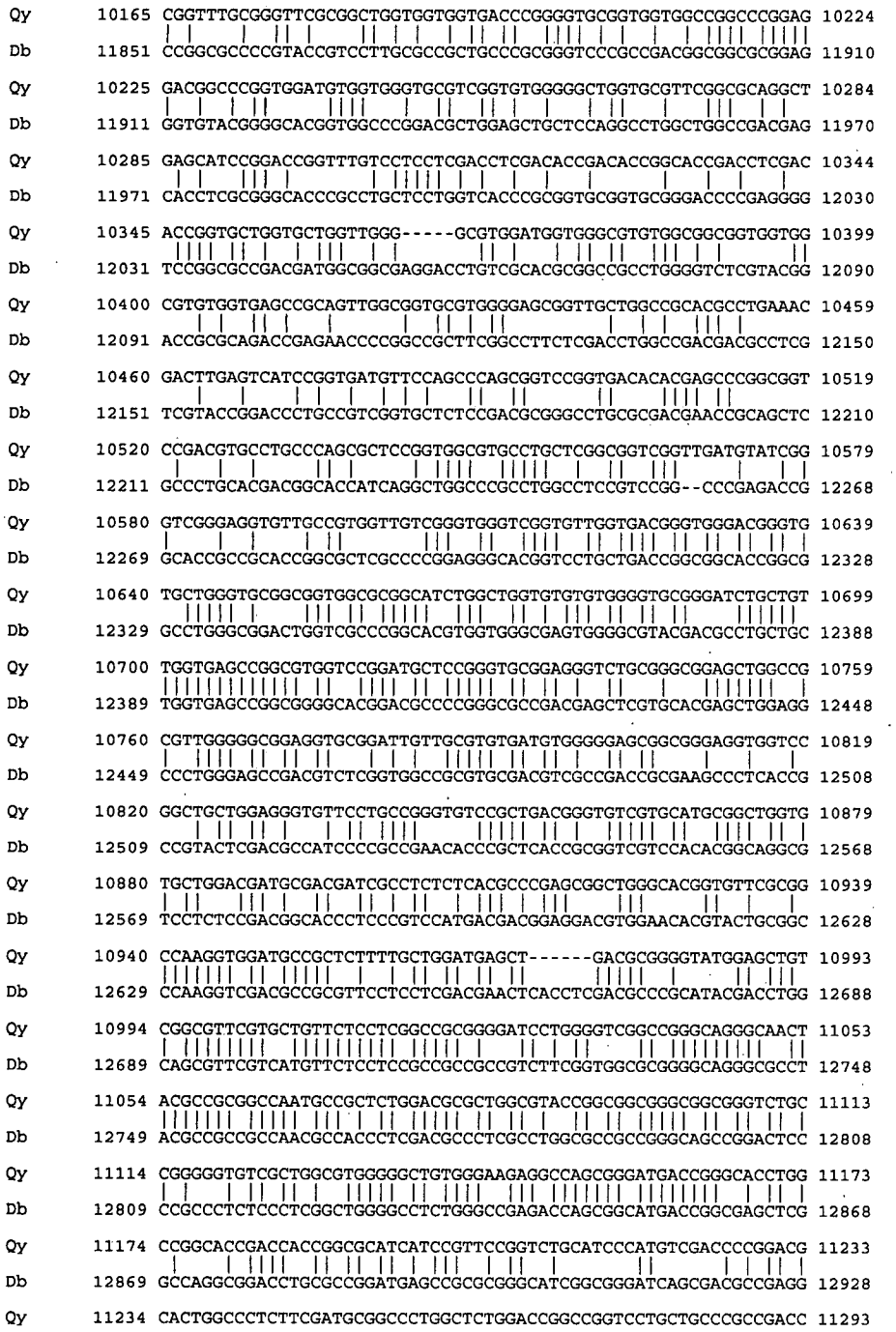


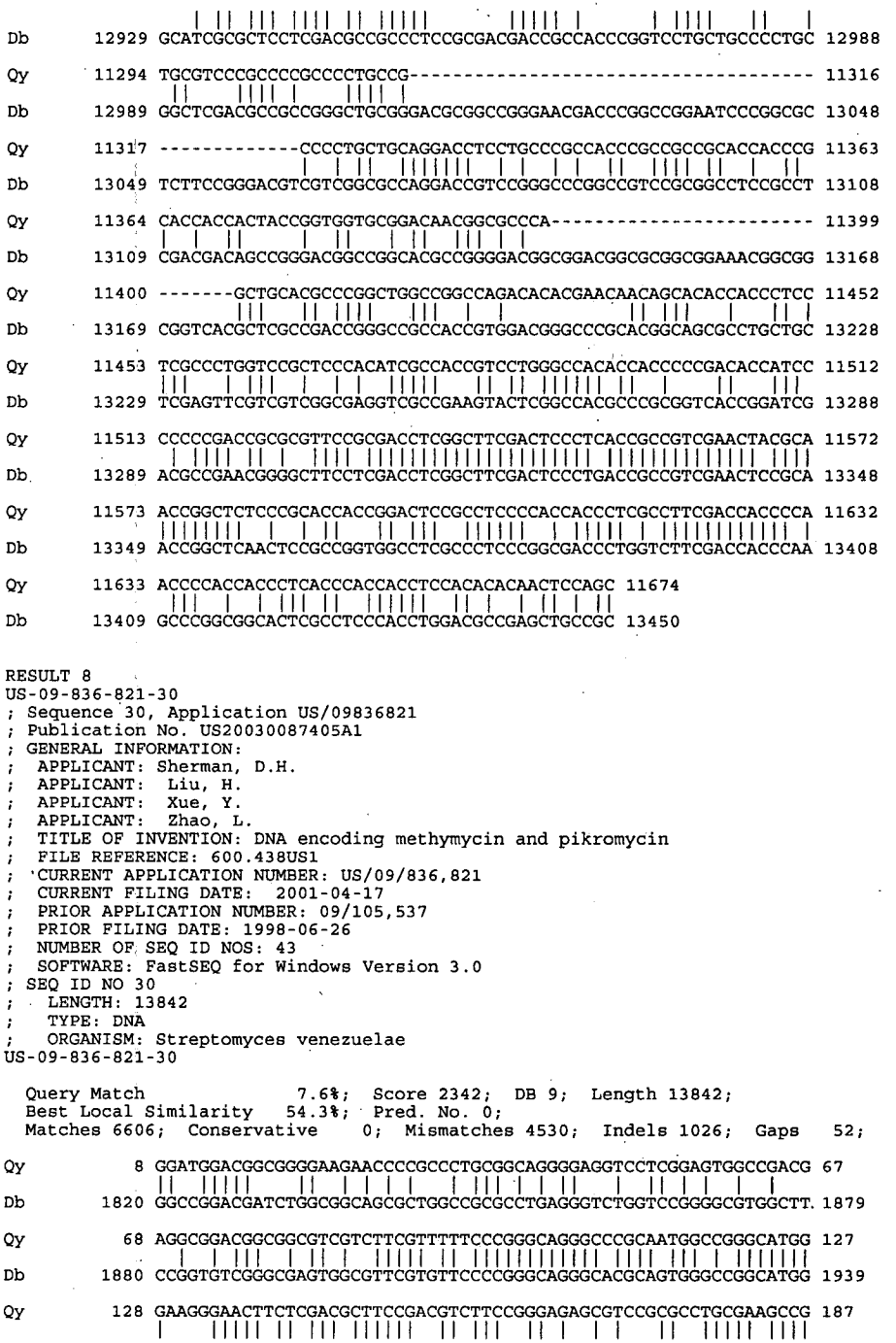

Qy 6550 TGGGAAACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGA 6609
Db 8104 TGGGAGGCCTTCGAGCGGGCCGGCATCGAGCCGGCATCGCTGCGCGGCAGCAGCACCGGT 8163
Qy 6610 GTCTTCACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCA 6669
Db 8164 GTCTTCATCGGCCTCTCCTACCAGGACTACGCGGCCCGCGTCCCGAACGCCCCGCGTGGC 8223
Qy 6670 ACCGATGGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTAC 6729
Db 8224 GTGGAGGGTTACCTGCTGACCGGCAGCACGCCGAGCGTCGCGTCGGGCCGTATCGCGTAC 8283
Qy 6730 ACGTTTGGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTG 6789
Db 8284 ACCTTCGGTCTCGAAGGGCCCGCGACGACCGTCGACACCGCCTGCTCGTCGTCGCTGACC 8343
Qy 6790 GCTTTGCATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGG 6849
Db 8344 GCCCTGCACCTGGCGGTGCGGGCGCTGCGCAGCGGCGAGTGCACGATGGCGCTCGCCGGT 8403
Qy 6850 GGTGTGACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTG 6909
Db 8404 GGCGTGGCGATGATGGCGACCCCGCACATGTTCGTGGAGTTCAGCCGTCAGCGGGCGCTC 8463
Qy 6910 GCCGCGGACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAG 6969
Db 8464 GCCCCGGACGGCCGCAGCAAGGCCTTCTCGGCGGACGCCGACGGGTTCGGCGCCGCGGAG 8523
Qy 6970 GGTGTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTC 7029
Db 8524 GGCGTCGGCCTGCTGCTCGTGGAGCGGCTCTCGGACGCGCGGCGCAACGGTCACCCGGTG 8583
Qy 7030 CTGGCCGTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCG 7089
Db 8584 CTCGCCGTGGTCCGCGGTACCGCCGTCAACCAGGACGGCGCCAGCAACGGGCTGACCGCG 8643
Qy 7090 CCCAACGGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCG 7149
Db 8644 CCCAACGGACCCTCGCAGCAGCGGGTGATCCGGCAGGCGCTCGCCGACGCCCGGCTGGCA 8703
Qy 7150 GCCGGTGATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATC 7209
Db 8704 CCCGGCGACATCGACGCCGTCGAGACGCACGGCACGGGAACCTCGCTGGGCGACCCCATC 8763
Qy 7210 GAGGCCCAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGG 7269
Db 8764 GAGGCCCAGGGCCTCCAGGCCACGTACGGCAAGGAGCGGCCCGCGGAACGGCCGCTCGCC 8823
Qy 7270 CTGGGCTCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTG 7329
Db 8824 ATCGGCTCCGTGAAGTCCAACATCGGACACACCCAGGCCGCGGCCGGTGCGGCGGGCATC 8883
Qy 7330 ATCAAGATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAG 7389
Db 8884 ATCAAGATGGTCCTCGCGATGCGCCACGGCACCCTGCCGAAGACCCTCCACGCCGACGAG 8943
Qy 7390 CCGTCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCC 7449
Db 8944 CCGAGCCCGCACGTCGACTGGGCGAACAGCGGCCTGGCCCTCGTCACCGAGCCGATCGAC 9003
Qy 7450 TGGCCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGC 7509
Db 9004 TGGCCGGCCGGCACCGGTCCG---CGCCGCGCCGCCGTCTCCTCCTTCGGCATCAGCGGG 9060
Qy 7510 ACCAACGCCCACGTCATCCTCGAAGAAGCACCCGCCGACGACGTT---------- 7554
Db 9061 ACGAACGCGCACGTCGTGCTGGAGCAGGCGCCGGATGCTGCTGGTGAGGTGCTTGGGGCC 9120
Qy 7555 ----------CCGGGGGGACCACCC 7569
Db 9121 GATGAGGTGCCTGAGGTGTCTGAGACGGTAGCGATGGCTGGGACGGCTGGGACCTCCGAG 9180
Qy 7570 GCCGGCGAGGGTGACGCGGGCAGCGACGATGAGGCTGCTGCCGGCAGTCCTGGGGTGT-- 7627
Db 9181 GTCGCTGAGGGCTCTGAGGCCTCCGAGGCCCCCGCGGCCCCCGGCAGCCGTGAGGCGTCC 9240
Qy 7628 ----------GGCCGTGGCTGGTGTCGGCCAAGTCGCAGCCGGCCCTGCGCGCCCAG 7674
Db 9241 CTCCCCGGGCACCTGCCCTGGGTGCTGTCCGCCAAGGACGAGCAGTCGCTGCGCGGCCAG 9300
Qy 7675 GCCCAGGCCCTGCACGCCCACCTCACCGACCACCCCGGC---------- 7713
Db 9301 GCCGCCGCCCTGCACGCGTGGCTGTCCGAGCCCGCCGCCGACCTGTCGGACGCGGACGGA 9360
Qy 7714 ---CTCGACCTCGCGGATGTCGGATACACCCTCGCCCACGCCCGCGCCGTGTTCGACCAC 7770
Db 9361 CCGGCCCGCCTGCGGGACGTCGGGTACACGCTCGCCACGAGCCGTACCGCCTTCGCGCAC 9420
Qy 7771 CGCGCCACCCTCATCGCCGCGGACCGCGACACGTTCCTGCAAGCACTCCAGGCACTCGCC 7830
Db 9421 CGCGCCGCCGTGACCGCCGCCGACCGGGACGGGTTCCTGGACGGGCTGGCCACGCTGGCC 9480
Qy 7831 GCAGGCGAGCCCCACCCCGCCGTCATCCACAGCAGCGCCCCGGGCGGGACCGGGACCGGG 7890
Db 9481 CAGGGCGGCACCTCGGCCCACGTCCACCTGGACACCGCCCGGGACGGCACC--------- 9531
Qy 7891 GAGGCCGCAGGAAAGACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATG 7950
Db 9532 ----------ACCGCGTTCCTCTTCACCGGCCAGGGCAGTCAGCGCCCCGGCGCC 9576
Qy 7951 GCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACC 8010
Db 9577 GGCCGTGAGCTGTACGACCGGCACCCCGTCTTCGCCCGGGCGCTCGACGAGATCTGCGCC 9636
Qy 8011 CACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGAC 8070
Db 9637 CACCTCGACGGTCACCTCGAACTGCCCCTGCTCGACGTGAT------GTTCGCGGCCGAG 9690
Qy 8071 AACGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCC 8130
Db 9691 GGCAGCGCGGAGGCCGCGCTGCTCGACGAGACGCGGTACACGCAGTGCGCGCTGTTCGCC 9750
Qy 8131 TTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTAC 8190
Db 9751 CTGGAGGTCGCGCTCTTCCGGCTCGT---CGAGAGCTGGGGCATGCGGCCGGCCGCACTG 9807
Qy 8191 GCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACC 8250
Db 9808 CTCGGTCACTCGGTCGGCGAGATCGCCGCCGCGCACGTCGCCGGTGTGTTCTCGCTCGCC 9867
Qy 8251 GACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCC---GGC 8307
Db 9868 GACGCCGCCCGCCTGGTCGCCGCGCGCGGCCGGCTCATGCAGGAGCTGCCCGCCGGTGGC 9927
Qy 8308 ACCATGACCACCCTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAA 8367
Db 9928 GCGATGCTCGCCGTCCAGGCCGCGGAGGACGAGATCCGCGTGTGGCTGGAGACGGAGGAG 9987
Qy 8368 AACGAC---------CTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGC 8418
Db 9988 CGGTACGCGGGACGTCTGGACGTCGCCGCCGTCAACGGCCCCGAGGCCGCCGTCCTGTCC 10047
Qy 8419 GGCACCCCCCACACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACC 8478
Db 10048 GGCGACGCGGACGCGGCGCGGGAGGCGGAGGCGTACTGGTCCGGGCTCGGCCGCAGGACC 10107
Qy 8479 AAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAA 8538
Db 10108 CGCGCGCTGCGGGTCAGCCACGCCTTCCACTCCGCGCACATGGACGGCATGCTCGACGGG 10167
Qy 8539 CTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCC 8598
Db 10168 TTCCGCGCCGTCCTGGAGACGGTGGAGTTCCGGCGCCCCTCCCTGACCGTGGTCTCGAAC 10227
Qy 8599 AACA----------CCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAA 8646


Db 6115 GCCGCCGTGCTCACCGCGCTGGTCGACGCCGGGGCGAAGGTCGAGGTGCTGACGGCCGGG 6174
Qy 4468 ACCTACCAGGAAACCTGGAAACCCCTCACCCTCCCCACCACCCACCAACCCCACCAAACC 4527
Db 6175 GCGGACGACGACCGTGAGGCCCTCGCCGCCCGGCTCACCGCACTGACG---------- 6222
Qy 4528 TGGCTCATCGCCATCCCCGAAACCCAGACCCACCACCCCCACATCACCAACATCCTCACC 4587
Db 6223 ---------- 6222
Qy 4588 AACCTCCACCACCACGGCATCACCCCCATCCCCCTCACCCTCAACCACACCCACACCAAC 4647
Db 6223 ---------- 6222
Qy 4648 CCCCAACACCTCCACCACACCCTCCACCACACCCGACAACAAGCCCAAAACCACACCACC 4707
Db 6223 ----------ACCGGT 6228
Qy 4708 GGAGCCATCACCGGCCTGCTCTCCCTCCTCGCCCTCGACGAAACACCCCACCCCCACCAC 4767
Db 6229 GACGGCTTCACCGGCGTGGTCTCGCTCCTCG---------- 6259
Qy 4768 CCCCACACACCCACCGGCACCCTCCTCAACCTCACCCTCACCCAAACCCACACCCAAACC 4827
Db 6260 ------------ACGGACTCGTACCGCAGGTCGCCTGGGTCCAGGCGCTCGGCGACGCC 6306
Qy 4828 CACCCACCAACCCCCCTCTGGTACGCCACCACCAACGCCACCACCACCCACCCCAACGAC 4887
Db 6307 GGAATCAAGGCGCCCCTGTGGTCCGTCACCCAGGGCGCGGTCTCCGTCGGACGTCTCGAC 6366
Qy 4888 CCCCTCACACACCCCACCCAAGCCCAAACCTGGGGACTCGCCCGCACCACCCTCCTCGAA 4947
Db 6367 ACCCCCGCCGACCCCGACCGGGCCATGCTCTGGGGCCTCGGCCGCGTCGTCGCCCTTGAG 6426
Qy 4948 CACCCCACCCACACCGCCGGAATCATCGACCTCCCCACCACCCCCACCCCCCACACCCTC 5007
Db 6427 CACCCCGAACGCTGGGCCGGCCTCGTCGACCTCCCCGCCCAGCCCGATGCCGCCGCCCTC 6486
Qy 5008 CAGCACCTCACCCAAACCCTCACCCAACCCCACCACCAAACCCAACTCGCCATCCGCACC 5067
Db 6487 GCCCACCTCGTCACCGCACTCTCCGGCGCCACCGGCGAGGACCAGATCGCCATCCGCACC 6546
Qy 5068 ACCGGCACCCACACCCGCCGCCTCACCCCCACCACCCTCACCCCCACACACCAACCACCC 5127
Db 6547 ACCGGACTCCACGCCCGCCGCCTCGCCCGCGCACCCCTCCACGGACGTCGGCCCACCCGC 6606
Qy 5128 ACCCCCACCCCCCACGGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCAC 5187
Db 6607 GACTGGCAGCCCCACGGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCAC 6666
Qy 5188 CTCACCCACCACCTCACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACC 5247
Db 6667 ---GCCGCACGCTGGATGGCCCACCACGGAGCCGAACACCTCCTCCTCGTCAGCCGCAGC 6723
Qy 5248 GGCCCCCACACCCCCCACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCAC 5307
Db 6724 GGCGAACAAGCCCCCGGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGC 6783
Qy 5308 CTCACCATCACCACCTGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACC 5367
Db 6784 GTCACCATCGCCGCCTGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCC 6843
Qy 5368 ATCCCCCCACAACACCCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCC 5427
Db 6844 ATCCCCGCCGAGACGCCCCTCACCGCCGTCGTCCACACCGCCGGCGCGCTCGACGACGGC 6903
Qy 5428 ACCCTCACCAACCTCACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGC 5487
Db 6904 ATCGTGGACACGCTGACCGCCGAGCAGGTCCGGCGGGCCCACCGTGCGAAGGCCGTCGGC 6963
Qy 5488 GCCCACCTCCTCCACCAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCC 5547
Db 6964 GCCTCGGTGCTCGACGAGCTGACCCGGGACCTCGACCTCGACGCGTTCGTGCTCTTCTCG 7023
Qy 5548 TCCGCCGCCGCCACCTTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTAC 5607
Db 7024 TCCGTGTCGAGCACTCTGGGCATCCCCGGTCAGGGCAACTACGCCCCGCACAACGCCTAC 7083
Qy 5608 CTCGACGCCCTCGCCCACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGG 5667
Db 7084 CTCGACGCCCTCGCGGCTCGCCGCCGGGCCACCGGCCGGTCCGCCGTCTCGGTGGCCTGG 7143
Qy 5668 GGCACCTGGCAAGGAAACGGACTCGCTGATTCGGACAAGGCCCGCGCATATCTCGACCGC 5727
Db 7144 GGACCGTGGGACGGTGGCGGCATGGCCGCCGGTGACGGCGTGGCCGAGCGGCTGCGCAAC 7203
Qy 5728 CGCGGGTTTCGACCCATGTCACCCGAGTTGGCCACGGCAGC---------- 5768
Db 7204 CACGGCGTGCCCGGCATGGACCCGGAACTCGCCCTGGCCGCACTGGAGTCCGCGCTCGGC 7263
Qy 5769 --------GGTCACGCAGGCGATCGCGGACACCGAACGGCCGTATGTCGTCATCGCCGAC 5820
Db 7264 CGGGACGAGACCGCGATCACCGTCGCGGACATCGACTGGGACCGCTTCTACCTCGCGTAC 7323
Qy 5821 ATCGACTGGAGCAAGATCGAACACACCTCTCAGACCAGCGACCTG------GTGAGCGCG 5874
Db 7324 TCCTCCGGTCGCCCGCAGCCCCTCGTCGAGGAGCTGCCCGAGGTGCGGCGCATCATCGAC 7383
Qy 5875 GCCCGGGAAAGGGAGCCAGCTGTCCAGCGCCCCACTCCACCGGCGG---------AGTTG 5925
Db 7384 GCACGGGACAGCGCCACGTCCGGACAGGGCGGGAGCTCCGCCCAGGGCGCCAACCCCCTG 7443
Qy 5926 CACAAAACGCTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTA 5985
Db 7444 GCCGAGCGGCTGGCCGCCGCGGCTCCCGGCGAGCGTACGGAGATCCTCCTCGGTCTCGTA 7503
Qy 5986 CGAGACCATGTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAG 6045
Db 7504 CGGGCGCAGGCCGCCGCCGTCGCTCCGGATGCGTTCGCCGGAGGACGTCGCCGCCGACCGC 7563
Qy 6046 TCGTTCCGTGCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATC 6105
Db 7564 GCCTTCAAGGACATCGGCTTCGACTCGCTCGCCGGTGTCGAGCTGCGCAACAGGCTGACC 7623
Qy 6106 AAGGCAACAGGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAA 6165
Db 7624 CGGGCGACCGGGCTCCAGCTGCCCGCGACGCTCGTCTTCGACCACCCGACGCCGCTGGCC 7683
Qy 6166 CTCGCCGTACACCTGCAGAACCAACTGCGGGGCACAGCAGCGGAGTCG---------- 6213
Db 7684 CTCGTGTCGCTGCTCCGCAGCGAGTTCCTCGGTGACGAGGAGACGGCGGACGCCCGGCGG 7743
Qy 6214 ----------GCTCCTTCAGCGGCAGCCGTTACCGCCGAGGCTTCT 6249
Db 7744 TCCGCGGCGCTGCCCGCGACTGTCGGTGCCGGTGCCGGCGCCGGCGCCGGCACCGATGCC 7803
Qy 6250 GTCACCGAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCG 6309
Db 7804 GACGACGATCCGATCGCGATCGTCGCGATGAGCTGCCGCTACCCCGGTGACATCCGCAGC 7863
Qy 6310 GCGGACGACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACC 6369
Db 7864 CCGGAGGACCTGTGGCGGATGCTGTCCGAGGGCGGCGAGGGCATCACGCCGTTCCCCACC 7923
Qy 6370 GACCGCGGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGC 6429
Db 7924 GACCGCGGCTGGGACCTCGACGGCCTGTACGACGCCGACCCGGACGCGCTCGGCAGGGCG 7983
Qy 6430 TACACCCGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGC 6489
Db 7984 TACGTCCGCGAGGGCGGGTTCCTGCACGACGCGGCCGAGTTCGACGCGGAGTTCTTCGGC 8043
Qy 6490 ATCAGCCCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCC 6549
Db 8044 GTCTCGCCGCGCGAGGCGCTGGCCATGGACCCGCAGCAGCGGATGCTCCTGACGACGTCC 8103


Qy 2254 AACCAGGACGGTGCGTCGAATGGGCTGACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTG 2313
Db 4126 AACCAGGACGGCGCGAGCAACGGCCTCACGGCTCCGAACGGGCCCTCGCAGCAGCGCGTC 4185
Qy 2314 ATCCGGCAGGCGTTGGCGAACGCGGGGTTGTCCGTGGCGGATGTGGATGTGGTGGAGGGG 2373
Db 4186 ATCCGGCGCGCGCTGGCGGACGCCCGGCTGACGACCTCCGACGTGGACGTCGTCGAGGCA 4245
Qy 2374 CACGGGACGGGCACGACGCTGGGTGATCCGATCGAGGCACAGGCGTTGCTCGCCACGTAC 2433
Db 4246 CACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCGCAGGCCCTGATCGCCACCTAC 4305
Qy 2434 GGGCAG---CGGGCCGGTGACAGGCCGCTGTGGCTGGGGTCTCTGAAGTCCAACATCGGG 2490
Db 4306 GGCCAGGGCCGTGACGACGAACAGCCGCTGCGCCTCGGGTCGTTGAAGTCCAACATCGGG 4365
Qy 2491 CACACCATGGCTGCCGCGGGTGTGGGTGGGGTCATCAAGATGGTGATGGCGTTGCGGGAG 2550
Db 4366 CACACCCAGGCCGCGGCCGGCGTCTCCGGTGTCATCAAGATGGTCCAGGCGATGCGCCAC 4425
Qy 2551 GGGGTGTTGCCGCGGACGTTGCATGTGGATAAGCCGTCGCCGCAGGTGGACTGGTCCGCG 2610
Db 4426 GGACTGCTGCCGAAGACGCTGCACGTCGACGAGCCCTCGGACCAGATCGACTGGTCGGCT 4485
Qy 2611 GGGGCGGTGCGGCTGCTGACGGAGGCGGTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTG 2670
Db 4486 GGCGCCGTGGAACTCCTCACCGAGGCCGTCGACTGGCCGGAGAAGCAGGACGGCGGGCTG 4545
Qy 2671 CGGCGGGCGGGAGTGTCGTCGTTCGGGATCGGCGGCACGAATGCGCATGTGATTTTGGAG 2730
Db 4546 CGCCGGGCCGCCGTCTCCTCCTTCGGGATCAGCGGCACCAATGCGCATGTGGTGCTCGAA 4605
Qy 2731 GAGGCGCCGGCGGCGGGGGGCTGTGTTGCCGGGGGTGGGGTGTTGGAGGGTGCTCCGGGT 2790
Db 4606 GAGGCCCCGGTGGTTGTCGAGGGTGCTTCGGTCGTCGAGCCGTCGGTTGGCGGGTCGGCG 4665
Qy 2791 CTTGCCATTTCGGTGGCTGAGTCGGTGGCCGCTCCAGTGGCTGTGTCTGCGCCGGTGGCT 2850
Db 4666 GTCGGCGGCGGTGTGACGCCTTGGGTGGTGTCGGCGAAGTCCGCTGCCGCGCTCGACGCG 4725
Qy 2851 GAGTCGGTGCCGGTGCCGGTGCCGGTGCCGGTTCCTGTGCCGGTGTCGGCTAGGTCTGAG 2910
Db 4726 CAGATCGAGCGGCT-----TGCCGCATTCGCCTCGCGGGATCGTACGGATGACGCCGACG 4780
Qy 2911 GCTGGGTTGCGGGCGCAGGCGGAGGCGTTGCGTCAGTACGTGGCAGTCCGGCCGGACGTT 2970
Db 4781 CCGGTGCTGTCGACGCGGG-- 4799
Qy 2971 TCGCTTGCCGATGTGGGTGCGGGTCTGGCCTGTGGGCGGGCTGTGCTGGAGCATCGTGCG 3030
Db 4800 --------CGCTGTCGCTCACGTACTGGCTGACGGGCGTGCTCAGTTCGAGCACCGGGCC 4851
Qy 3031 GTCGTCCTGGCCGCGGACCGTGAGGAGCTGGTGCAAGGGTTGGGGGCGCTGGCGGCGGGT 3090
Db 4852 GTCGCGCTCGGCGCCGGGGCGGACGACCTCGTACAGGCGCTGGCCGATCCGGACGGGC-- 4909
Qy 3091 GAGCCGGATCGGCGGGTGACCACGGGTCATGCGCCGGGTGGTGACCGGGGCGGTGTCGTC 3150
Db 4910 ----------------------TGATACGCGGAACGGCTTCCGGTGTCGGGCGAGTGGCG 4947
Qy 3151 TTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCC 3210
Db 4948 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCTGGCATGGGTGCCGAACTGCTGGACTCT 5007
Qy 3211 TCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGAC 3270
Db 5008 TCCGCGGTGTTCGCGGCGGCCATGGCCGAGTGTGAGGCCGCGCTGTCCCCGTACGTCGAC 5067
Qy 3271 TGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGAT 3330
Db 5068 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 5127
Qy 3331 GTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTAC 3390
Db 5128 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 5187
Qy 3391 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 3450
Db 5188 GGTGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 5247
Qy 3451 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 3510
Db 5248 GCCGGAGCCCTGCCCCTGGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 5307
Qy 3511 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 3567
Db 5308 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAACGAGGACGCCGTC 5367
Qy 3568 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 3627
Db 5368 CTGGAGCGACTGAGTGA------CTTCGACGGGCTGTCCGTCGCCGCCGTCAACGGGCCC 5421
Qy 3628 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 3687
Db 5422 ACCGCCACTGTCGTGTCGGGTGACCCCGTACAGATCGAAGAGCTTGCTCAGGCGTGCAAG 5481
Qy 3688 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 3747
Db 5482 GCGGACGGATTCCGCGCGCGGATCATTCCCGTCGACTACGCGTCCCACAGCCGGCAGGTC 5541
Qy 3748 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 3807
Db 5542 GAGATCATCGAGAGCGAGCTCGCCCAGGTCCTCGCCGGTCTCAGCCCGCAGGCCCCGCGC 5601
Qy 3808 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 3867
Db 5602 GTGCCGTTCTTCTCGACGCTCGAAGGCACCTGGATCACCGAGCCCGTCCTCGACGGCACC 5661
Qy 3868 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 3927
Db 5662 TACTGGTACCGCAACCTCCGTCACCGCGTCGGCTTCGCCCCCGCCATCGAGACCCTGGC- 5720
Qy 3928 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 3987
Db 5721 ----------------CGTCGACGAGGGCTTCACGCACTTCGTCGAGGTCAGCGCCCAC 5763
Qy 3988 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 4047
Db 5764 CCCGTCCTCACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGCACCCTCCGTCGCGAA 5823
Qy 4048 GACAACGACACCCGCCGCTTCCTCACCGCCCTCGCCCACACCCATACCACCGGCATCGGC 4107
Db 5824 CAGGGAGGCCAAGAGCGTCTGGTCACCTCGCTCGCCGAGGCGTGGGTCAACGGGCTTCCC 5883
Qy 4108 ACACCCACCACCTGGCACCACCACTACACCCACCACCACACCCACCCCCACCCCCACACG 4167
Db 5884 GTGGCATGGACTTCGC----------TCCTGCCCGCCACGGCCTCC 5919
Qy 4168 CACCTCGACCTGCCCCACCTACCCCTTCCAACACCAGCACTACTGGCTCGAGAGCTCACAG 4227
Db 5920 CGCCCCGGTCTGCCCACCTACGCCTTCCAGGCCGAGCGCTACTGGCTCGAGAACACTCCC 5979
Qy 4228 CCGGGTGCCGGATCCGGTTCGGGTGCCGGTGCCGGTTCGGGTGCCGGTTCCGGGCGGGCA 4287
Db 5980 GCCG----------CCCTGGCCACCG 5995
Qy 4288 GGGACTGCGGGCGGGACGGCAGAGGTGGAGTCGCGGTTCTGGGACGCGGTGGCCCGCCAG 4347
Db 5996 GCGACGACTGGCGCTACCGCATCGACTG-GAAGCGCCTCCCGGCCGCCGAGGGGTCCGAG 6054
Qy 4348 GACCTGGAAACGGTCGCGACCACACTCGCCGTGCCCCCCTCCGCCGGCCTGGACACGGTG 4407
Db 6055 CGCACCGGCCTGTCCGGCCGCTGGCTCGCCGTCACGCCGGAGGACCACTCCGCGCAGGCC 6114
Qy 4408 GTGCCCGCACTCTCCGCCTGGCACCGCCACCAACACGACCAAGCCCGCATCAACACCTGG 4467

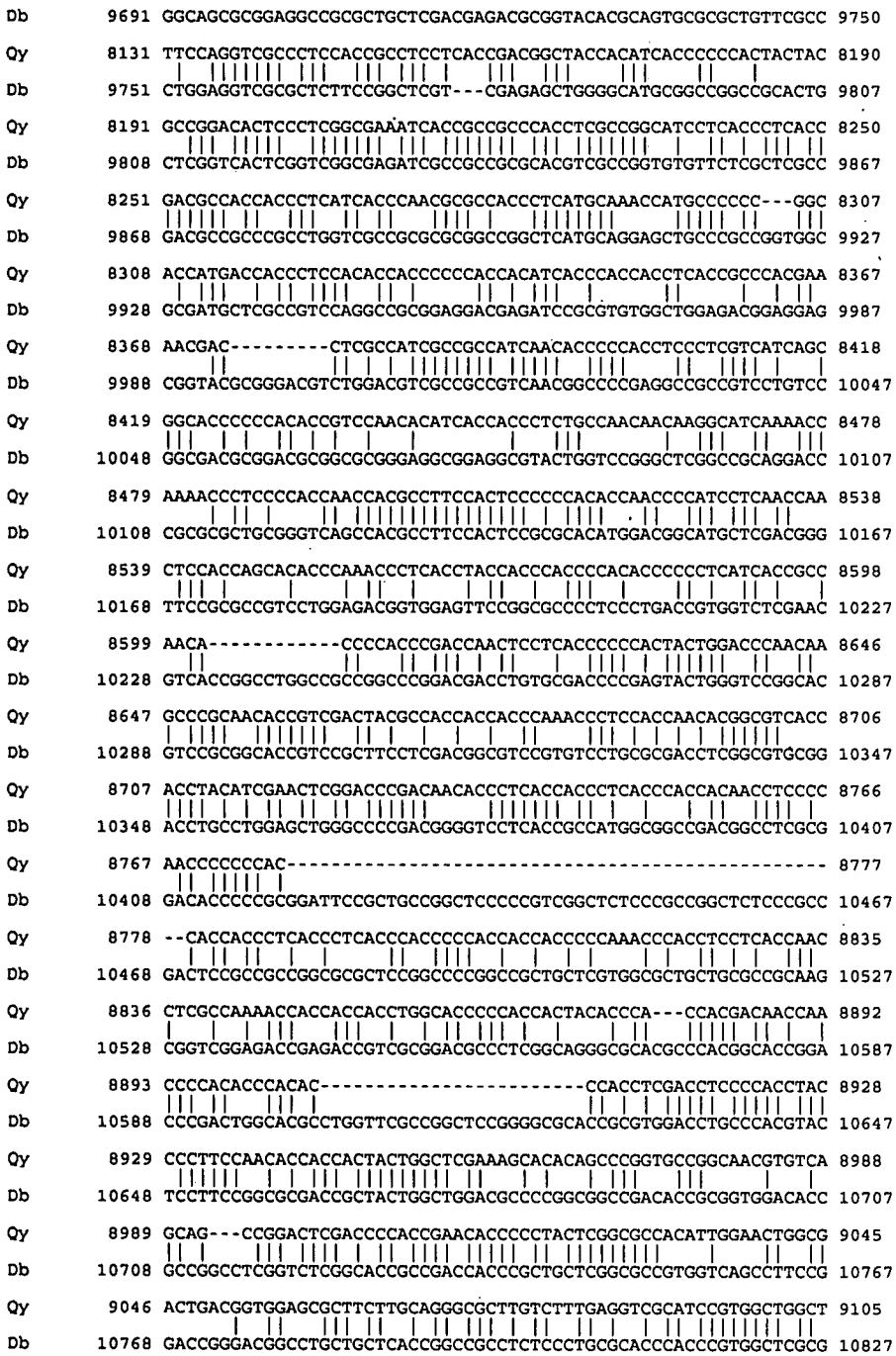


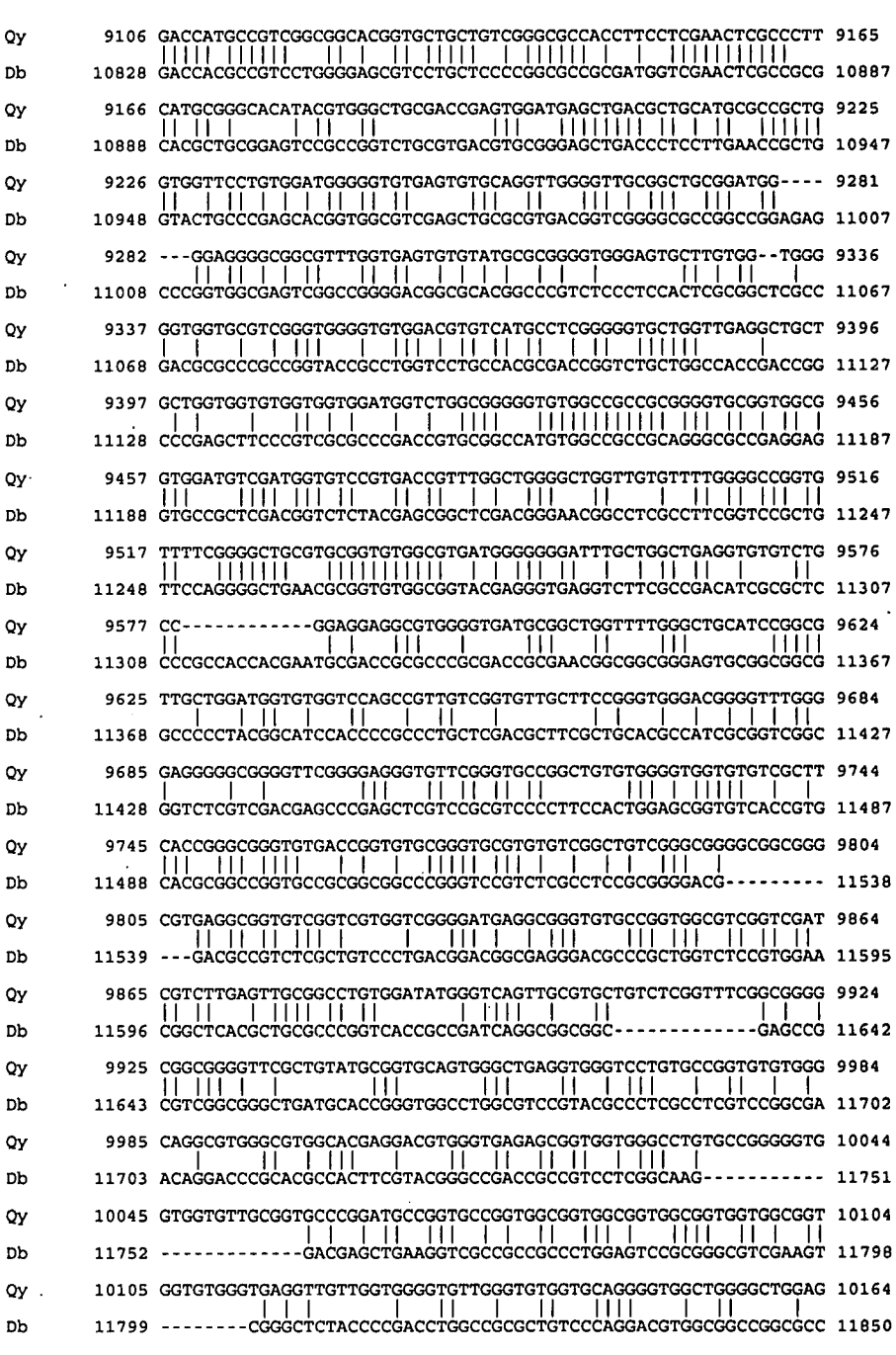


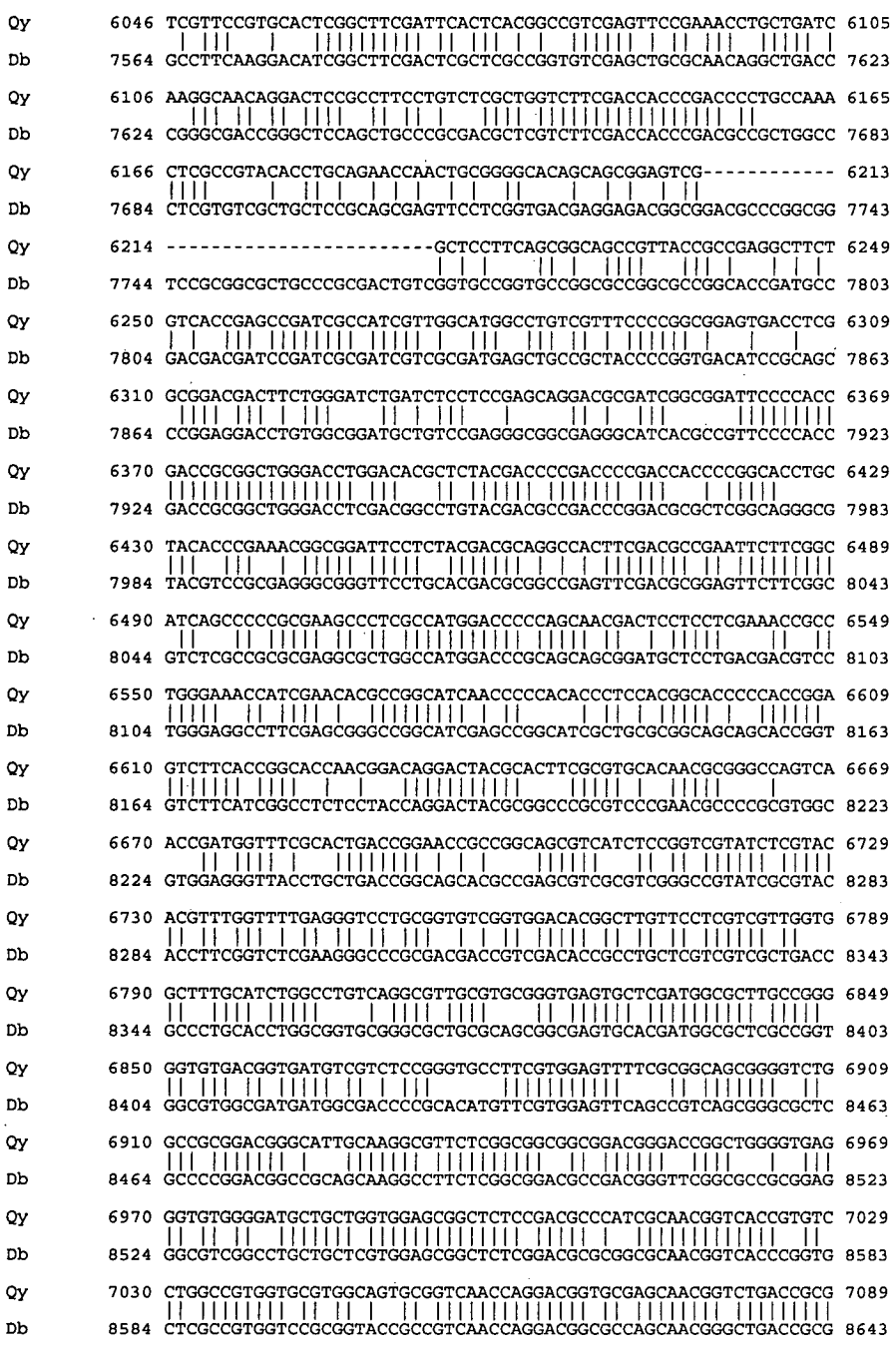


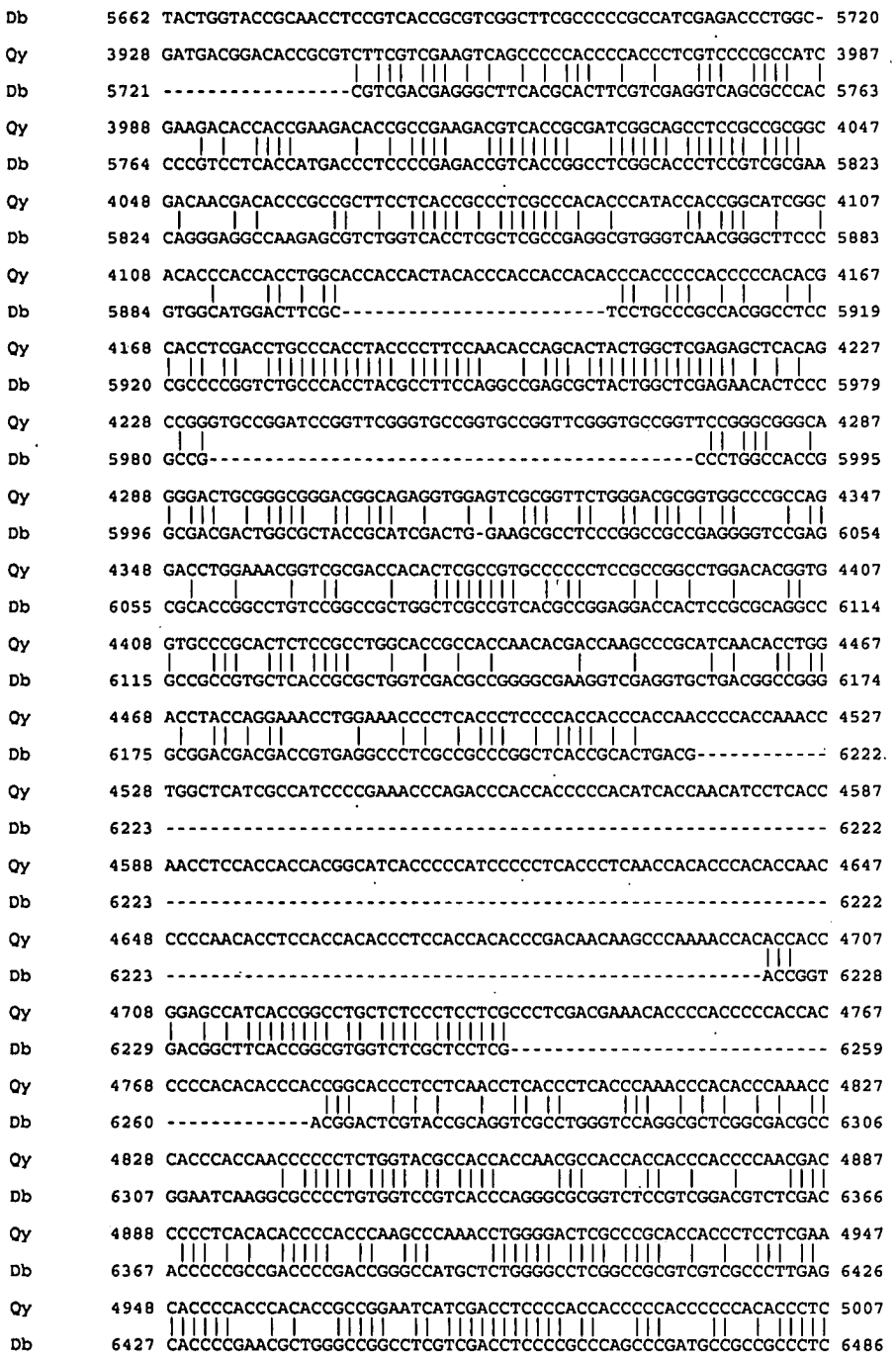
RESULT 8
US-09-836-821-30
; Sequence 30, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-836-821-30

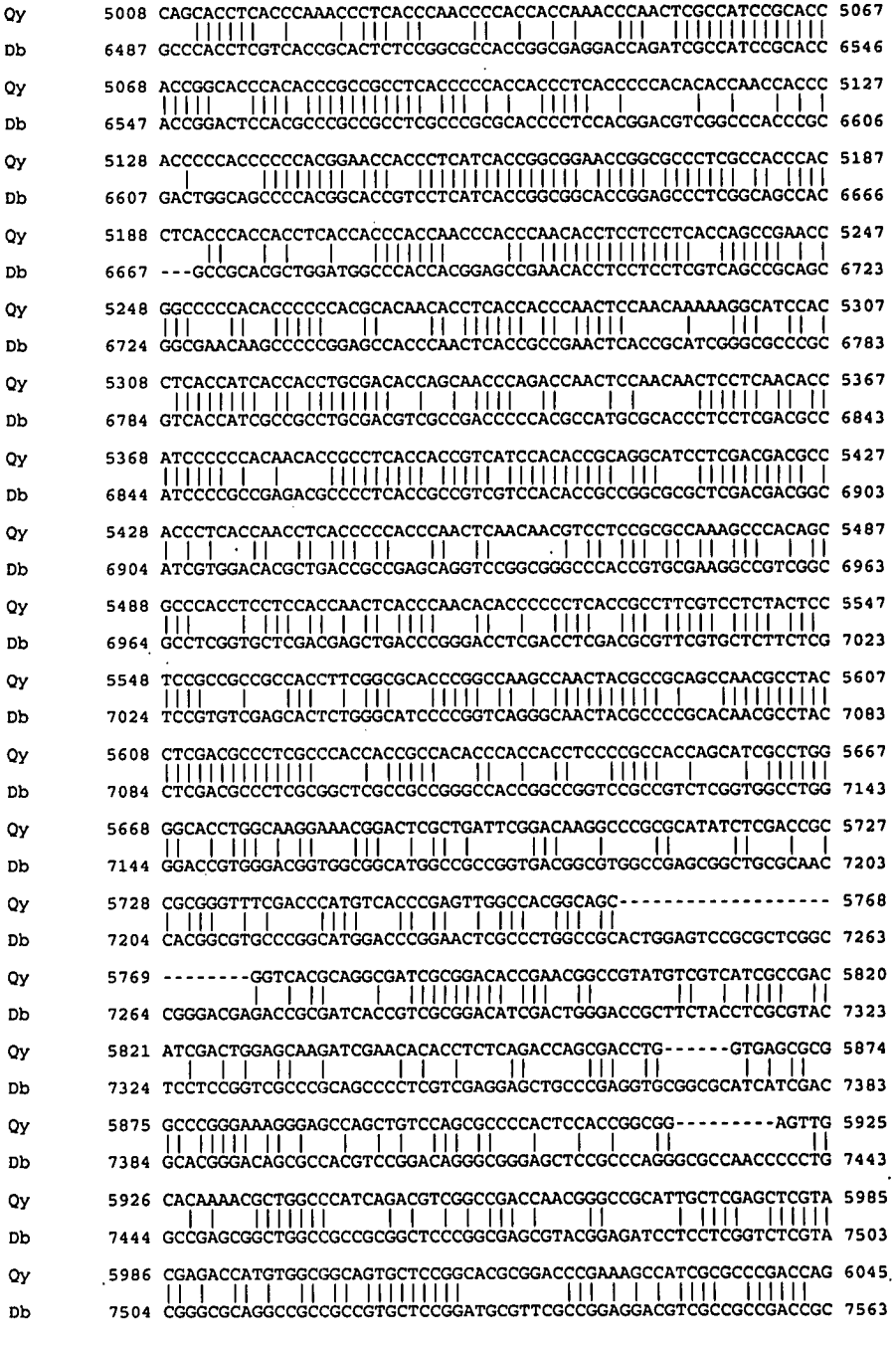
Query Match 7.6%; Score 2342; DB 9; Length 13842;
Best Local Similarity 54.3%; Pred. No. 0;
Matches 6606; Conservative 0; Mismatches 4530; Indels 1026; Gaps 52;

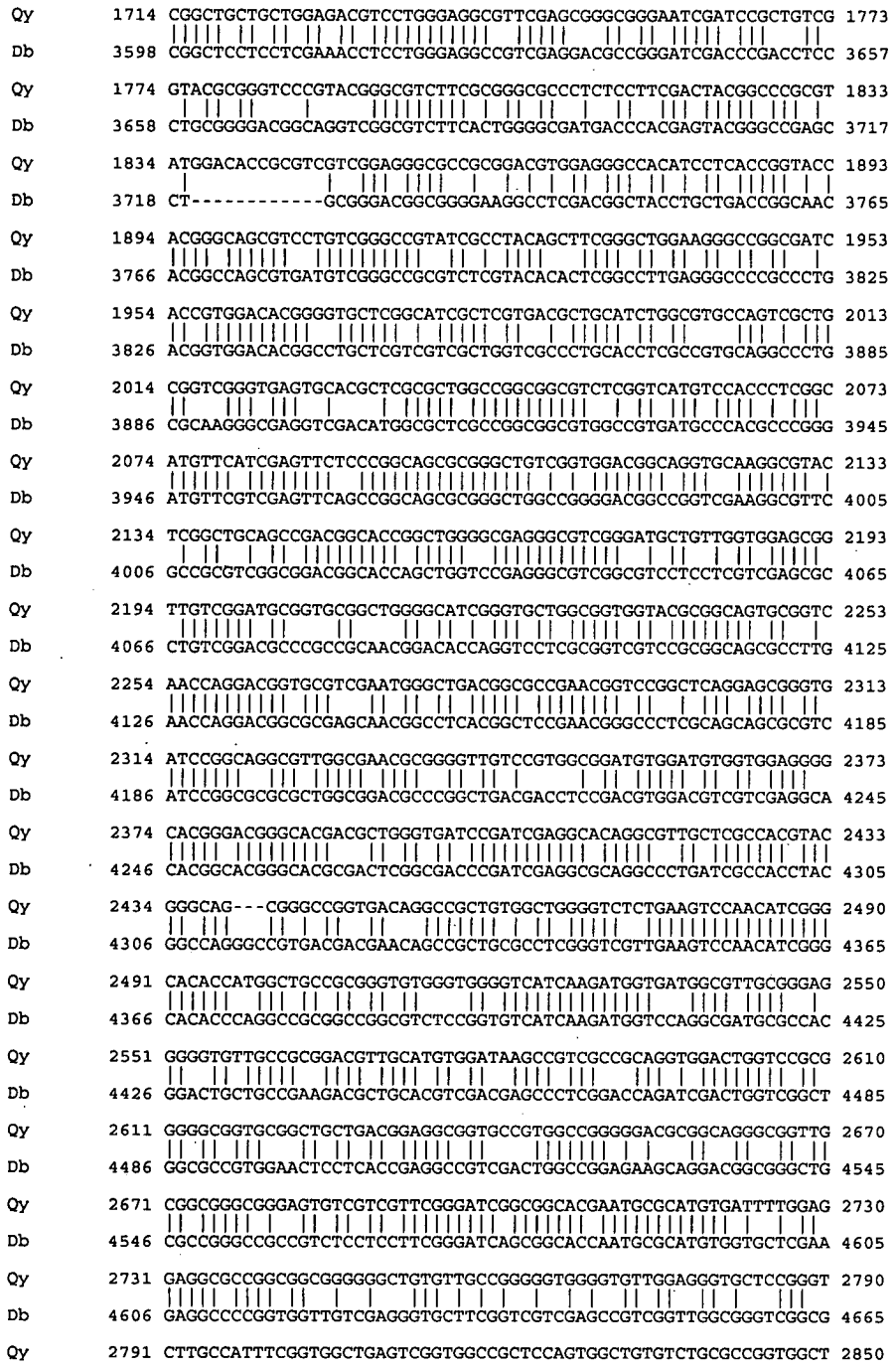





Qy 6046 TCGTTCCGTGCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATC 6105
Db 7564 GCCTTCAAGGACATCGGCTTCGACTCGCTCGCCGGTGTCGAGCTGCGCAACAGGCTGACC 7623
Qy 6106 AAGGCAACAGGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAA 6165
Db 7624 CGGGCGACCGGGCTCCAGCTGCCCGCGACGCTCGTCTTCGACCACCCGACGCCGCTGGCC 7683
Qy 6166 CTCGCCGTACACCTGCAGAACCAACTGCGGGGCACAGCAGCGGAGTCG---------- 6213
Db 7684 CTCGTGTCGCTGCTCCGCAGCGAGTTCCTCGGTGACGAGGAGACGGCGGACGCCCGGCGG 7743
Qy 6214 ----------GCTCCTTCAGCGGCAGCCGTTACCGCCGAGGCTTCT 6249
Db 7744 TCCGCGGCGCTGCCCGCGACTGTCGGTGCCGGTGCCGGCGCCGGCGCCGGCACCGATGCC 7803
Qy 6250 GTCACCGAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCG 6309
Db 7804 GACGACGATCCGATCGCGATCGTCGCGATGAGCTGCCGCTACCCCGGTGACATCCGCAGC 7863
Qy 6310 GCGGACGACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACC 6369
Db 7864 CCGGAGGACCTGTGGCGGATGCTGTCCGAGGGCGGCGAGGGCATCACGCCGTTCCCCACC 7923
Qy 6370 GACCGCGGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGC 6429
Db 7924 GACCGCGGCTGGGACCTCGACGGCCTGTACGACGCCGACCCGGACGCGCTCGGCAGGGCG 7983
Qy 6430 TACACCCGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGC 6489
Db 7984 TACGTCCGCGAGGGCGGGTTCCTGCACGACGCGGCCGAGTTCGACGCGGAGTTCTTCGGC 8043
Qy 6490 ATCAGCCCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCC 6549
Db 8044 GTCTCGCCGCGCGAGGCGCTGGCCATGGACCCGCAGCAGCGGATGCTCCTGACGACGTCC 8103
Qy 6550 TGGGAAACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGA 6609
Db 8104 TGGGAGGCCTTCGAGCGGGCCGGCATCGAGCCGGCATCGCTGCGCGGCAGCAGCACCGGT 8163
Qy 6610 GTCTTCACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCA 6669
Db 8164 GTCTTCATCGGCCTCTCCTACCAGGACTACGCGGCCCGCGTCCCGAACGCCCCGCGTGGC 8223
Qy 6670 ACCGATGGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTAC 6729
Db 8224 GTGGAGGGTTACCTGCTGACCGGCAGCACGCCGAGCGTCGCGTCGGGCCGTATCGCGTAC 8283
Qy 6730 ACGTTTGGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTG 6789
Db 8284 ACCTTCGGTCTCGAAGGGCCCGCGACGACCGTCGACACCGCCTGCTCGTCGTCGCTGACC 8343
Qy 6790 GCTTTGCATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGG 6849
Db 8344 GCCCTGCACCTGGCGGTGCGGGCGCTGCGCAGCGGCGAGTGCACGATGGCGCTCGCCGGT 8403
Qy 6850 GGTGTGACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTG 6909
Db 8404 GGCGTGGCGATGATGGCGACCCCGCACATGTTCGTGGAGTTCAGCCGTCAGCGGGCGCTC 8463
Qy 6910 GCCGCGGACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAG 6969
Db 8464 GCCCCGGACGGCCGCAGCAAGGCCTTCTCGGCGGACGCCGACGGGTTCGGCGCCGCGGAG 8523
Qy 6970 GGTGTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTC 7029
Db 8524 GGCGTCGGCCTGCTGCTCGTGGAGCGGCTCTCGGACGCGCGGCGCAACGGTCACCCGGTG 8583
Qy 7030 CTGGCCGTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCG 7089
Db 8584 CTCGCCGTGGTCCGCGGTACCGCCGTCAACCAGGACGGCGCCAGCAACGGGCTGACCGCG 8643
Qy 7090 CCCAACGGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCG 7149
Db 8644 CCCAACGGACCCTCGCAGCAGCGGGTGATCCGGCAGGCGCTCGCCGACGCCCGGCTGGCA 8703
Qy 7150 GCCGGTGATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATC 7209
Db 8704 CCCGGCGACATCGACGCCGTCGAGACGCACGGCACGGGAACCTCGCTGGGCGACCCCATC 8763
Qy 7210 GAGGCCCAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGG 7269
Db 8764 GAGGCCCAGGGCCTCCAGGCCACGTACGGCAAGGAGCGGCCCGCGGAACGGCCGCTCGCC 8823
Qy 7270 CTGGGCTCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTG 7329
Db 8824 ATCGGCTCCGTGAAGTCCAACATCGGACACACCCAGGCCGCGGCCGGTGCGGCGGGCATC 8883
Qy 7330 ATCAAGATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAG 7389
Db 8884 ATCAAGATGGTCCTCGCGATGCGCCACGGCACCCTGCCGAAGACCCTCCACGCCGACGAG 8943
Qy 7390 CCGTCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCC 7449
Db 8944 CCGAGCCCGCACGTCGACTGGGCGAACAGCGGCCTGGCCCTCGTCACCGAGCCGATCGAC 9003
Qy 7450 TGGCCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGC 7509
Db 9004 TGGCCGGCCGGCACCGGTCCG---CGCCGCGCCGCCGTCTCCTCCTTCGGCATCAGCGGG 9060
Qy 7510 ACCAACGCCCACGTCATCCTCGAAGAAGCACCCGCCGACGACGTT---------- 7554
Db 9061 ACGAACGCGCACGTCGTGCTGGAGCAGGCGCCGGATGCTGCTGGTGAGGTGCTTGGGGCC 9120
Qy 7555 ----------CCGGGGGGACCACCC 7569
Db 9121 GATGAGGTGCCTGAGGTGTCTGAGACGGTAGCGATGGCTGGGACGGCTGGGACCTCCGAG 9180
Qy 7570 GCCGGCGAGGGTGACGCGGGCAGCGACGATGAGGCTGCTGCCGGCAGTCCTGGGGTGT-- 7627
Db 9181 GTCGCTGAGGGCTCTGAGGCCTCCGAGGCCCCCGCGGCCCCCGGCAGCCGTGAGGCGTCC 9240
Qy 7628 ----------GGCCGTGGCTGGTGTCGGCCAAGTCGCAGCCGGCCCTGCGCGCCCAG 7674
Db 9241 CTCCCCGGGCACCTGCCCTGGGTGCTGTCCGCCAAGGACGAGCAGTCGCTGCGCGGCCAG 9300
Qy 7675 GCCCAGGCCCTGCACGCCCACCTCACCGACCACCCCGGC---------- 7713
Db 9301 GCCGCCGCCCTGCACGCGTGGCTGTCCGAGCCCGCCGCCGACCTGTCGGACGCGGACGGA 9360
Qy 7714 ---CTCGACCTCGCGGATGTCGGATACACCCTCGCCCACGCCCGCGCCGTGTTCGACCAC 7770
Db 9361 CCGGCCCGCCTGCGGGACGTCGGGTACACGCTCGCCACGAGCCGTACCGCCTTCGCGCAC 9420
Qy 7771 CGCGCCACCCTCATCGCCGCGGACCGCGACACGTTCCTGCAAGCACTCCAGGCACTCGCC 7830
Db 9421 CGCGCCGCCGTGACCGCCGCCGACCGGGACGGGTTCCTGGACGGGCTGGCCACGCTGGCC 9480
Qy 7831 GCAGGCGAGCCCCACCCCGCCGTCATCCACAGCAGCGCCCCGGGCGGGACCGGGACCGGG 7890
Db 9481 CAGGGCGGCACCTCGGCCCACGTCCACCTGGACACCGCCCGGGACGGCACC---------- 9531
Qy 7891 GAGGCCGCAGGAAAGACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATG 7950
Db 9532 ----------ACCGCGTTCCTCTTCACCGGCCAGGGCAGTCAGCGCCCCGGCGCC 9576
Qy 7951 GCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACC 8010
Db 9577 GGCCGTGAGCTGTACGACCGGCACCCCGTCTTCGCCCGGGCGCTCGACGAGATCTGCGCC 9636
Qy 8011 CACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGAC 8070
Db 9637 CACCTCGACGGTCACCTCGAACTGCCCCTGCTCGACGTGAT------GTTCGCGGCCGAG 9690
Qy 8071 AACGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCC 8130






Qy 1714 CGGCTGCTGCTGGAGACGTCCTGGGAGGCGTTCGAGCGGGCGGGAATCGATCCGCTGTCG 1773
Db 3598 CGGCTCCTCCTCGAAACCTCCTGGGAGGCCGTCGAGGACGCCGGGATCGACCCGACCTCC 3657
Qy 1774 GTACGCGGGTCCCGTACGGGCGTCTTCGCGGGCGCCCTCTCCTTCGACTACGGCCCGCGT 1833
Db 3658 CTGCGGGGACGGCAGGTCGGCGTCTTCACTGGGGCGATGACCCACGAGTACGGGCCGAGC 3717
Qy 1834 ATGGACACCGCGTCGTCGGAGGGCGCCGCGGACGTGGAGGGCCACATCCTCACCGGTACC 1893
Db 3718 CT-----------GCGGGACGGCGGGGAAGGCCTCGACGGCTACCTGCTGACCGGCAAC 3765
Qy 1894 ACGGGCAGCGTCCTGTCGGGCCGTATCGCCTACAGCTTCGGGCTGGAAGGGCCGGCGATC 1953
Db 3766 ACGGCCAGCGTGATGTCGGGCCGCGTCTCGTACACACTCGGCCTTGAGGGCCCCGCCCTG 3825
Qy 1954 ACCGTGGACACGGGGTGCTCGGCATCGCTCGTGACGCTGCATCTGGCGTGCCAGTCGCTG 2013
Db 3826 ACGGTGGACACGGCCTGCTCGTCGTCGCTGGTCGCCCTGCACCTCGCCGTGCAGGCCCTG 3885
Qy 2014 CGGTCGGGTGAGTGCACGCTCGCGCTGGCCGGCGGCGTCTCGGTCATGTCCACCCTCGGC 2073
Db 3886 CGCAAGGGCGAGGTCGACATGGCGCTCGCCGGCGGCGTGGCCGTGATGCCCACGCCCGGG 3945
Qy 2074 ATGTTCATCGAGTTCTCCCGGCAGCGCGGGCTGTCGGTGGACGGCAGGTGCAAGGCGTAC 2133
Db 3946 ATGTTCGTCGAGTTCAGCCGGCAGCGCGGGCTGGCCGGGGACGGCCGGTCGAAGGCGTTC 4005
Qy 2134 TCGGCTGCAGCCGACGGCACCGGCTGGGGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGG 2193
Db 4006 GCCGCGTCGGCGGACGGCACCAGCTGGTCCGAGGGCGTCGGCGTCCTCCTCGTCGAGCGC 4065
Qy 2194 TTGTCGGATGCGGTGCGGCTGGGGCATCGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTC 2253
Db 4066 CTGTCGGACGCCCGCCGCAACGGACACCAGGTCCTCGCGGTCGTCCGCGGCAGCGCCTTG 4125
Qy 2254 AACCAGGACGGTGCGTCGAATGGGCTGACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTG 2313
Db 4126 AACCAGGACGGCGCGAGCAACGGCCTCACGGCTCCGAACGGGCCCTCGCAGCAGCGCGTC 4185
Qy 2314 ATCCGGCAGGCGTTGGCGAACGCGGGGTTGTCCGTGGCGGATGTGGATGTGGTGGAGGGG 2373
Db 4186 ATCCGGCGCGCGCTGGCGGACGCCCGGCTGACGACCTCCGACGTGGACGTCGTCGAGGCA 4245
Qy 2374 CACGGGACGGGCACGACGCTGGGTGATCCGATCGAGGCACAGGCGTTGCTCGCCACGTAC 2433
Db 4246 CACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCGCAGGCCCTGATCGCCACCTAC 4305
Qy 2434 GGGCAG---CGGGCCGGTGACAGGCCGCTGTGGCTGGGGTCTCTGAAGTCCAACATCGGG 2490
Db 4306 GGCCAGGGCCGTGACGACGAACAGCCGCTGCGCCTCGGGTCGTTGAAGTCCAACATCGGG 4365
Qy 2491 CACACCATGGCTGCCGCGGGTGTGGGTGGGGTCATCAAGATGGTGATGGCGTTGCGGGAG 2550
Db 4366 CACACCCAGGCCGCGGCCGGCGTCTCCGGTGTCATCAAGATGGTCCAGGCGATGCGCCAC 4425
Qy 2551 GGGGTGTTGCCGCGGACGTTGCATGTGGATAAGCCGTCGCCGCAGGTGGACTGGTCCGCG 2610
Db 4426 GGACTGCTGCCGAAGACGCTGCACGTCGACGAGCCCTCGGACCAGATCGACTGGTCGGCT 4485
Qy 2611 GGGGCGGTGCGGCTGCTGACGGAGGCGGTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTG 2670
Db 4486 GGCGCCGTGGAACTCCTCACCGAGGCCGTCGACTGGCCGGAGAAGCAGGACGGCGGGCTG 4545
Qy 2671 CGGCGGGCGGGAGTGTCGTCGTTCGGGATCGGCGGCACGAATGCGCATGTGATTTTGGAG 2730
Db 4546 CGCCGGGCCGCCGTCTCCTCCTTCGGGATCAGCGGCACCAATGCGCATGTGGTGCTCGAA 4605
Qy 2731 GAGGCGCCGGCGGCGGGGGGCTGTGTTGCCGGGGGTGGGGTGTTGGAGGGTGCTCCGGGT 2790
Db 4606 GAGGCCCCGGTGGTTGTCGAGGGTGCTTCGGTCGTCGAGCCGTCGGTTGGCGGGTCGGCG 4665
Qy 2791 CTTGCCATTTCGGTGGCTGAGTCGGTGGCCGCTCCAGTGGCTGTGTCTGCGCCGGTGGCT 2850
Db 4666 GTCGGCGGCGGTGTGACGCCTTGGGTGGTGTCGGCGAAGTCCGCTGCCGCGCTCGACGCG 4725
Qy 2851 GAGTCGGTGCCGGTGCCGGTGCCGGTGCCGGTTCCTGTGCCGGTGTCGGCTAGGTCTGAG 2910
Db 4726 CAGATCGAGCGGCT-----TGCCGCATTCGCCTCGCGGGATCGTACGGATGACGCCGACG 4780
Qy 2911 GCTGGGTTGCGGGCGCAGGCGGAGGCGTTGCGTCAGTACGTGGCAGTCCGGCCGGACGTT 2970
Db 4781 CCGGTGCTGTCGACGCGGG-- 4799
Qy 2971 TCGCTTGCCGATGTGGGTGCGGGTCTGGCCTGTGGGCGGGCTGTGCTGGAGCATCGTGCG 3030
Db 4800 --------CGCTGTCGCTCACGTACTGGCTGACGGGCGTGCTCAGTTCGAGCACCGGGCC 4851
Qy 3031 GTCGTCCTGGCCGCGGACCGTGAGGAGCTGGTGCAAGGGTTGGGGGCGCTGGCGGCGGGT 3090
Db 4852 GTCGCGCTCGGCGCCGGGGCGGACGACCTCGTACAGGCGCTGGCCGATCCGGACGGGC-- 4909
Qy 3091 GAGCCGGATCGGCGGGTGACCACGGGTCATGCGCCGGGTGGTGACCGGGGCGGTGTCGTC 3150
Db 4910 ----------------------TGATACGCGGAACGGCTTCCGGTGTCGGGCGAGTGGCG 4947
Qy 3151 TTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCC 3210
Db 4948 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCTGGCATGGGTGCCGAACTGCTGGACTCT 5007
Qy 3211 TCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGAC 3270
Db 5008 TCCGCGGTGTTCGCGGCGGCCATGGCCGAGTGTGAGGCCGCGCTGTCCCCGTACGTCGAC 5067
Qy 3271 TGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGAT 3330
Db 5068 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 5127
Qy 3331 GTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTAC 3390
Db 5128 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 5187
Qy 3391 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 3450
Db 5188 GGTGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 5247
Qy 3451 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 3510
Db 5248 GCCGGAGCCCTGCCCCTGGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 5307
Qy 3511 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 3567
Db 5308 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAACGAGGACGCCGTC 5367
Qy 3568 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 3627
Db 5368 CTGGAGCGACTGAGTGA------CTTCGACGGGCTGTCCGTCGCCGCCGTCAACGGGCCC 5421
Qy 3628 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 3687
Db 5422 ACCGCCACTGTCGTGTCGGGTGACCCCGTACAGATCGAAGAGCTTGCTCAGGCGTGCAAG 5481
Qy 3688 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 3747
Db 5482 GCGGACGGATTCCGCGCGCGGATCATTCCCGTCGACTACGCGTCCCACAGCCGGCAGGTC 5541
Qy 3748 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 3807
Db 5542 GAGATCATCGAGAGCGAGCTCGCCCAGGTCCTCGCCGGTCTCAGCCCGCAGGCCCCGCGC 5601
Qy 3808 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 3867
Db 5602 GTGCCGTTCTTCTCGACGCTCGAAGGCACCTGGATCACCGAGCCCGTCCTCGACGGCACC 5661
Qy 3868 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 3927

; Sequence 30, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988,384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-30

Query Match 7.6%; Score 2342; DB 9; Length 13842;
Best Local Similarity 54.3%; Pred. No. 0;
Matches 6606; Conservative 0; Mismatches 4530; Indels 1026; Gaps 52;

```
Qy          8 GGATGGACGGCGGGGAAGAACCCCGCCCTGCGGCAGGGGAGGTCCTCGGAGTGGCCGACG 67
Db       1820 GGCCGGACGATCTGGCGGCAGCGCTGGCCGCGCCTGAGGGTCTGGTCCGGGGCGTGGCTT 1879

Qy         68 AGGCGGACGGCGGCGTCGTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGG 127
Db       1880 CCGGTGTCGGGCGAGTGGCGTTCGTGTTCCCCGGGCAGGGCACGCAGTGGGCCGGCATGG 1939

Qy        128 GAAGGGAACTTCTCGACGCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCG 187
Db       1940 GTGCCGAACTGCTGGACTCTTCCGCGGTGTTCGCGGCGGCCATGGCCGAATGCGAGGCCG 1999

Qy        188 CGTTCGCGCCCTACGTCGACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTC 247
Db       2000 CACTCTCCCCGTACGTCGACTGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGC 2059

Qy        248 CCGGGCTGGACCGGGTGGACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGG 307
Db       2060 CCACGCTGGAGCGGGTCGATGTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGG 2119

Qy        308 CCGCCCTCTGGCGCTCGCAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCG 367
Db       2120 CTCGCGTGTGGCAGCACCACGGGGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCG 2179

Qy        368 AGATCGCGGCAGCCCACGTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGA 427
Db       2180 AGATCGCCGCCGCGTACGTCGCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGA 2239

Qy        428 CGCTTTGGAGCCAGGCACAGACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCG 484
Db       2240 CCCTGCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCG 2299

Qy        485 CCGCCACGCCGGATGAGCTCCTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGC 544
Db       2300 CGCTGAGCGAGGACGCCGTCCTGGAGCGACTGGCCGGGTTCGACG--------------- 2344

Qy        545 GGCTCGCCGTCGCAGCCGTCAACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGG 604
Db       2345 GGCTGTCCGTCGCCGCTGTGAACGGGCCCACCGCCACCGTGGTCTCCGGTGACCCCGTAC 2404

Qy        605 CCGTCGCGGACCTGGTGGCCGACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGG 664
Db       2405 AGATCGAAGAGCTTGCTCGGGCGTGTGAGGCCGATGGGGTCCGTGCGCGGGTCATTCCCG 2464

Qy        665 TGGACGTTCCCGCCCACTCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCC 724
Db       2465 TCGACTACGCGTCCCACAGCCGGCAGGTCGAGATCATCGAGAGCGAGCTCGCCGAGGTCC 2524

Qy        725 TGCTGCCCATCACCCCACGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCC 784
Db       2525 TCGCCGGGCTCAGCCCGCAGGCTCCGCGCGTGCCGTTCTTCTCGACACTCGAAGGCGCCT 2584

Qy        785 GCCTCGACACCCGCGAGCTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCC 844
Db       2585 GGATCACCGAGCCCGTGCTCGACGGCGGCTACTGGTACCGCAACCTGCGCCATCGTGTGG 2644

Qy        845 GGTTCGAGCCCGCCGCCCGG---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGA 901
Db       2645 GCTTCGCCCCGGCCGTCGAGACCCTGGCCACCGACGAGGGCTTCACCCACTTCGTCGAGG 2704

Qy        902 TGAGCCCGCACCCGGTGCTGACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGG----- 956
Db       2705 TCAGCGCCCACCCCGTCCTCACCATGGCCCTCCCCGGGACCGTCACCGGTCTGGCGACCC 2764

Qy        957 ----------------------------------------------CGACACCACCGGCA 970
Db       2765 TGCGTCGCGACAACGGCGGTCAGGACCGCCTAGTCGCCTCCCTCGCCGAAGCATGGGCCA 2824

Qy        971 CCGCCGACACCGTGATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCC 1030
Db       2825 ACGGACTCGCGGTCGACTGGAGCCCGCTCCTCCCCTCCGCGACCGGCCACCACTCCGACC 2884

Qy       1031 TGACGTCTCTCGC----------------CCAACTACGGGGGCATGGTGAGACGTCGGCGA 1075
Db       2885 TCCCCACCTACGCGTTCCAGACCGAGCGCCACTGGCTGGGCGAGATCGAGGCGCTCGCCC 2944

Qy       1076 CCACCGTCCTCTCGGCACGCCTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCC 1135
Db       2945 CGGCGGGCGAGCCGGCGGTGCAGCCCGCCGTCCTCCGCACGGAGGCGGCCGAGCCGGCGG 3004

Qy       1136 TGGACCTGGTGCGCGCCCACACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCG 1195
Db       3005 AGCTCGACCGGGACGAGCAGCTGCGCGTGATCCTGGACAAGGTCCGGGCGCAGACGGCCC 3064

Qy       1196 CGTCGGATGCCGGCCCATCGG----------------------CGAGTTTCGCCCACCTCG 1234
Db       3065 AGGTGCTGGGGTACGCGACAGGCGGGCAGATCGAGGTCGACCGGACCTTCCGTGAGGCCG 3124

Qy       1235 GCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGC 1294
Db       3125 GTTGCACCTCCCTGACCGGCGTGGACCTGCGCAACCGGATCAACGCCGCCTTCGGCGTAC 3184

Qy       1295 GGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGCCGCGCGCCTTC 1354
Db       3185 GGATGGCGCCGTCCATGATCTTCGACTTCCCCACCCCCGAGGCTCTCGCGGAGCAGCTGC 3244

Qy       1355 GGACCGCGGCGCTCGGCCACCTCGACGAGGACACCGCGCCCGTACCGGACTCACCCAGCG 1414
Db       3245 TCCTCGTCGTGCACGG-------GGAGGCGGCGGCGAACCCGGCCGGTGCGGAGCCGGCT 3297

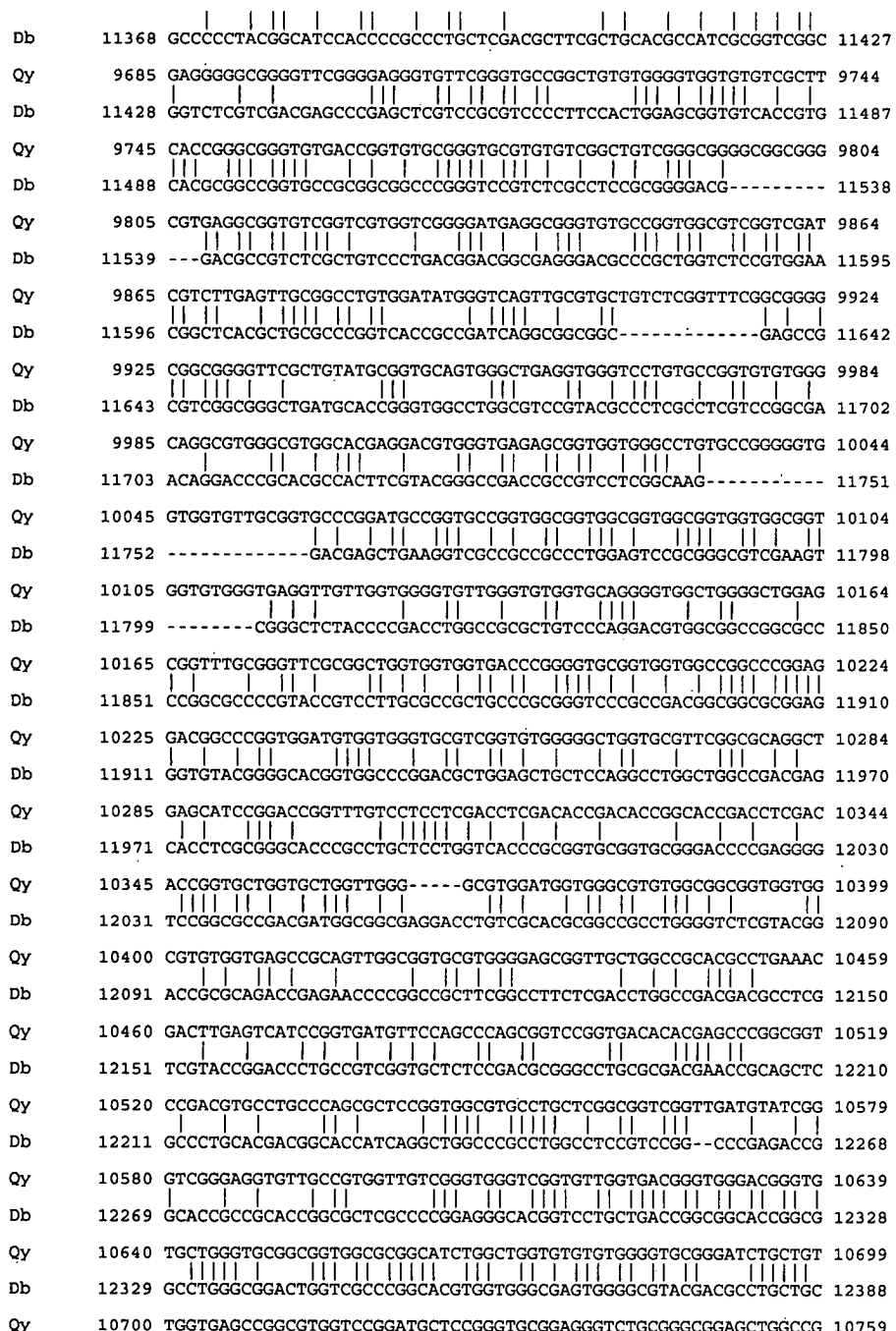
Qy       1415 GCCACGGAGGC-ACGGCAGCGGCGGACGACCCGATCGCCATCATCGGCATGGCATGCCGT 1473
Db       3298 CCGGTGGCGGCGGCCGGTGCCGTCGACGAGCCGGTGGCGATCGTCGGCATGGCCTGCCGC 3357

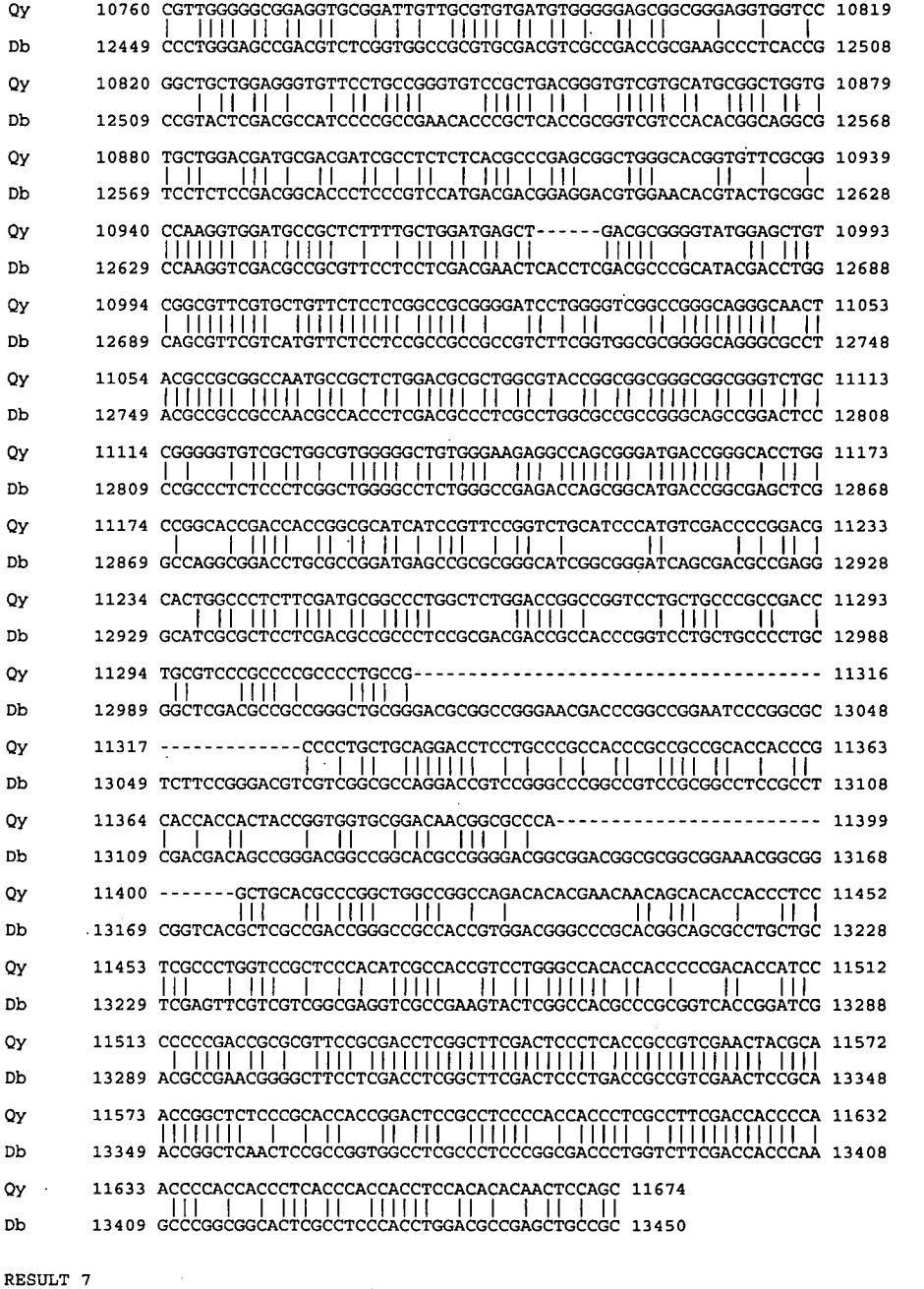
Qy       1474 TTCCCGGGCGGAGTCCGGTCCCCGAAGGACCTGTGGGAGCTGGCCGCCTCGGGCGGAGAC 1533
Db       3358 CTGCCCGGTGGGGTCGCCTCGCCGGAGGACCTGTGGCGGCTGGTGGCCGGCGGCGGGGAC 3417

Qy       1534 GCCATCGGGCCGTTCCCCACCGACCGCGGATGGCCCACGGAACAGCGTCACGCCCAGGAC 1593
Db       3418 GCGATCTCGGAGTTCCCGCAGGACCGCGGCTGGGACGTGGAGGGGCTGTACCACCCGGAT 3477

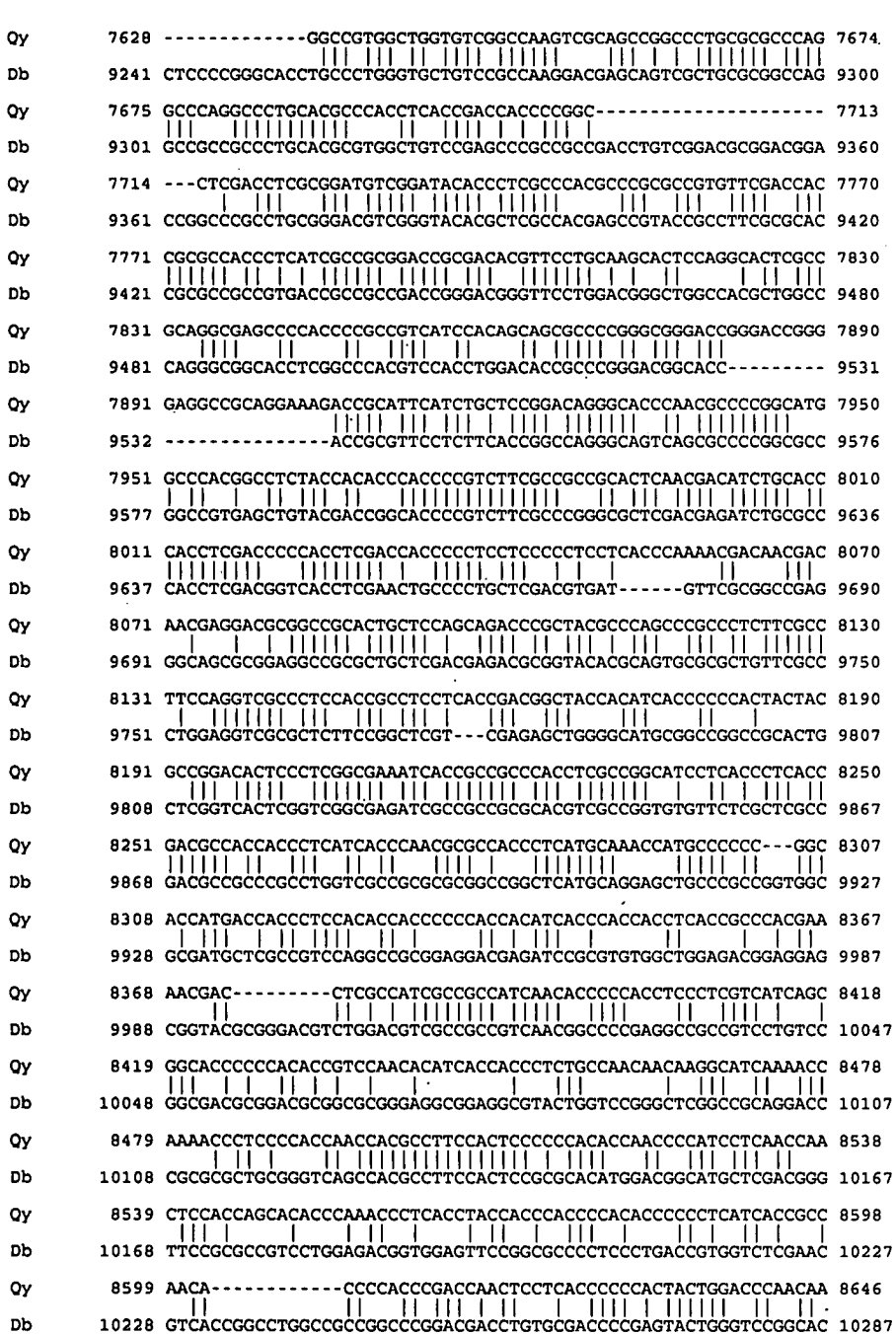
Qy       1594 CCCACGCAGCCCGGCACGTTCTATCCGCAGGGAGGCGGGTTCCTTCACGACGCGGCGCAC 1653
Db       3478 CCGGAGCACCCCGGCACGTCGTACGTCCGCCAGGGCGGTTTCATCGAGAACGTCGCCGGC 3537

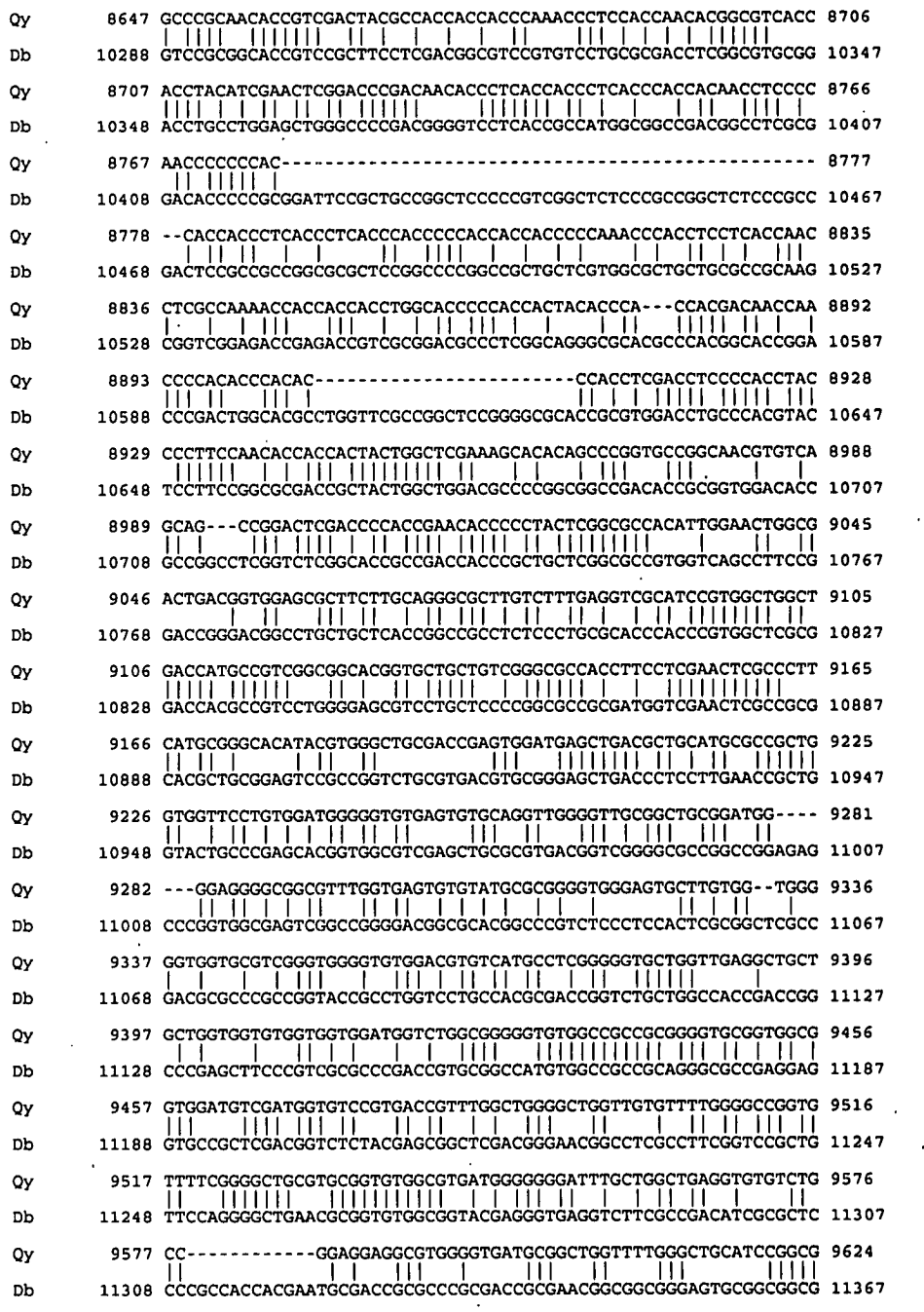
Qy       1654 TTCGACGCCGGCTTCTTCGGAATCAGTCCACGTGAGGCACTGGCGATGGATCCGCAGCAG 1713
Db       3538 TTCGACGCGGCCTTCTTCGGGATCTCGCCGCGCGAGGCCCTCGCCATGGACCCGCAGCAG 3597
```

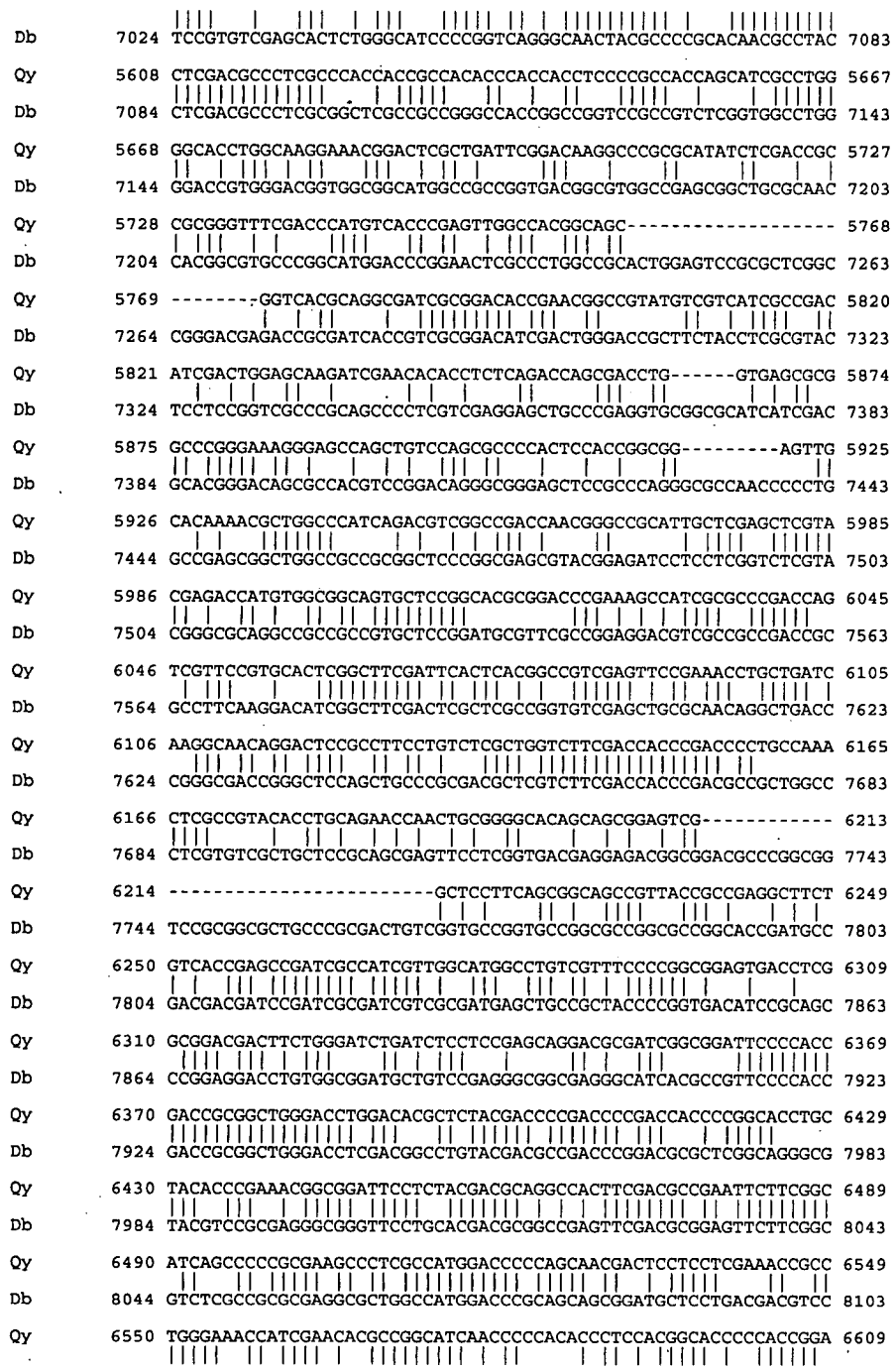


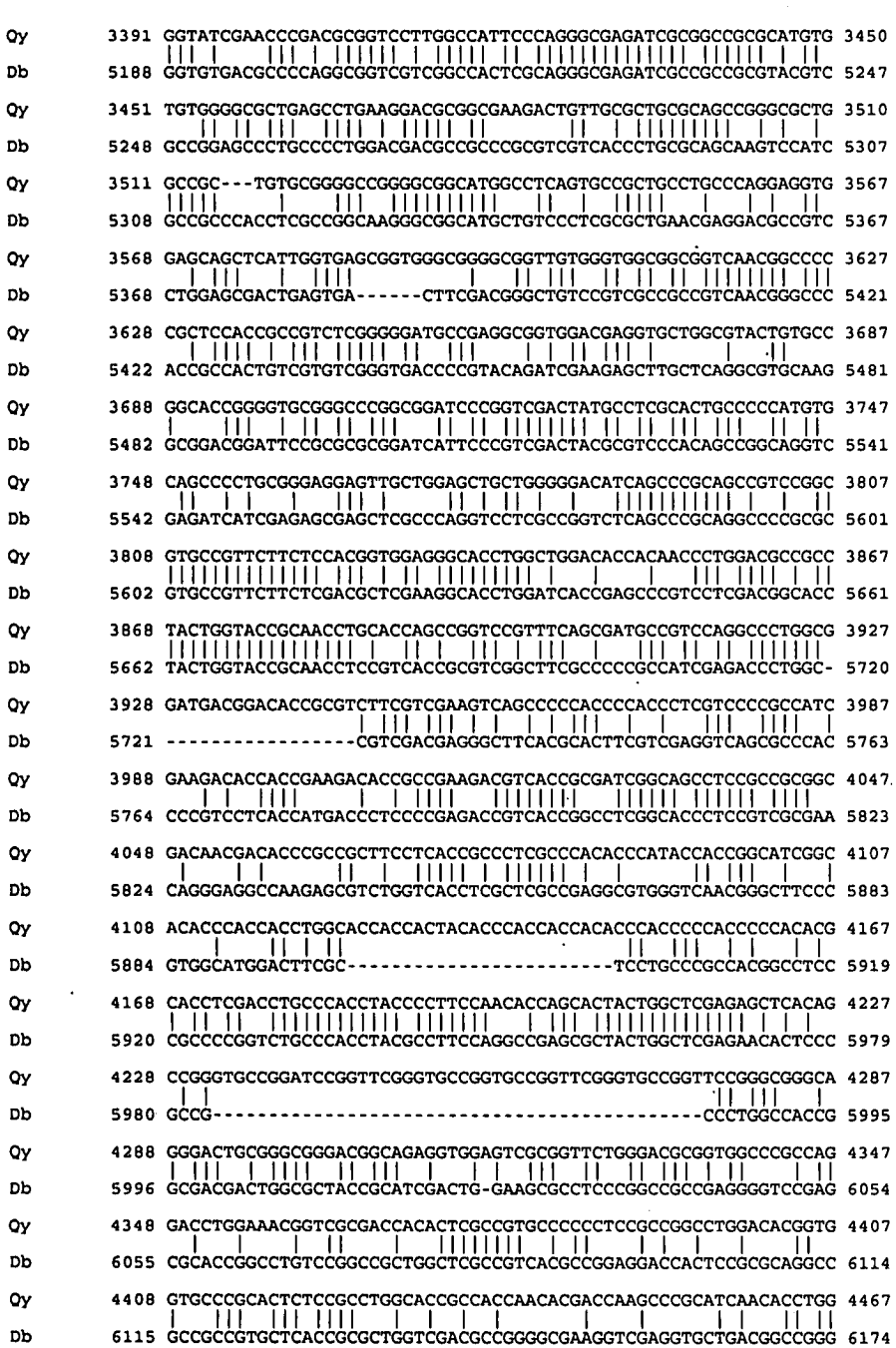


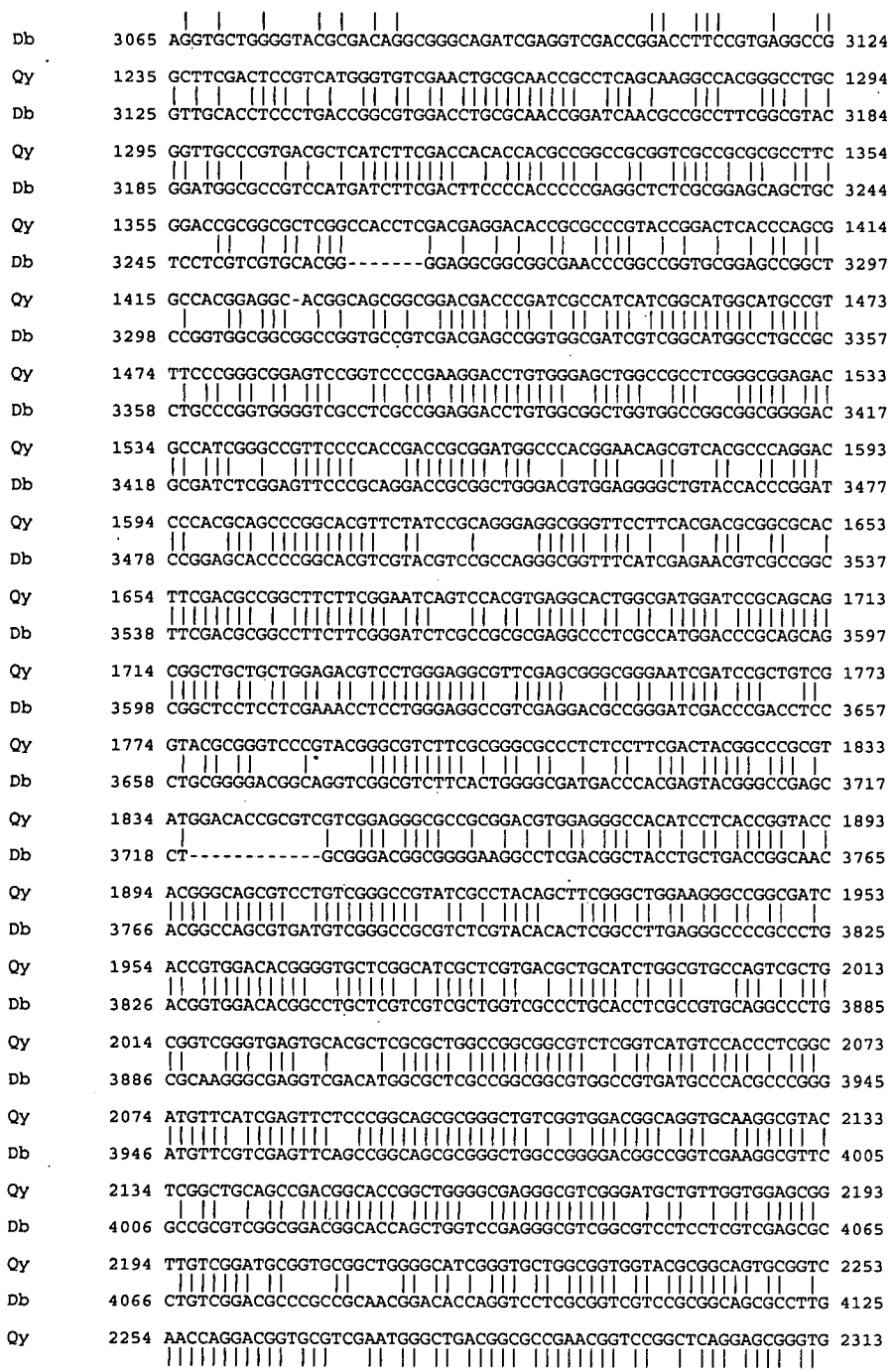
RESULT 7
US-09-988-384B-30

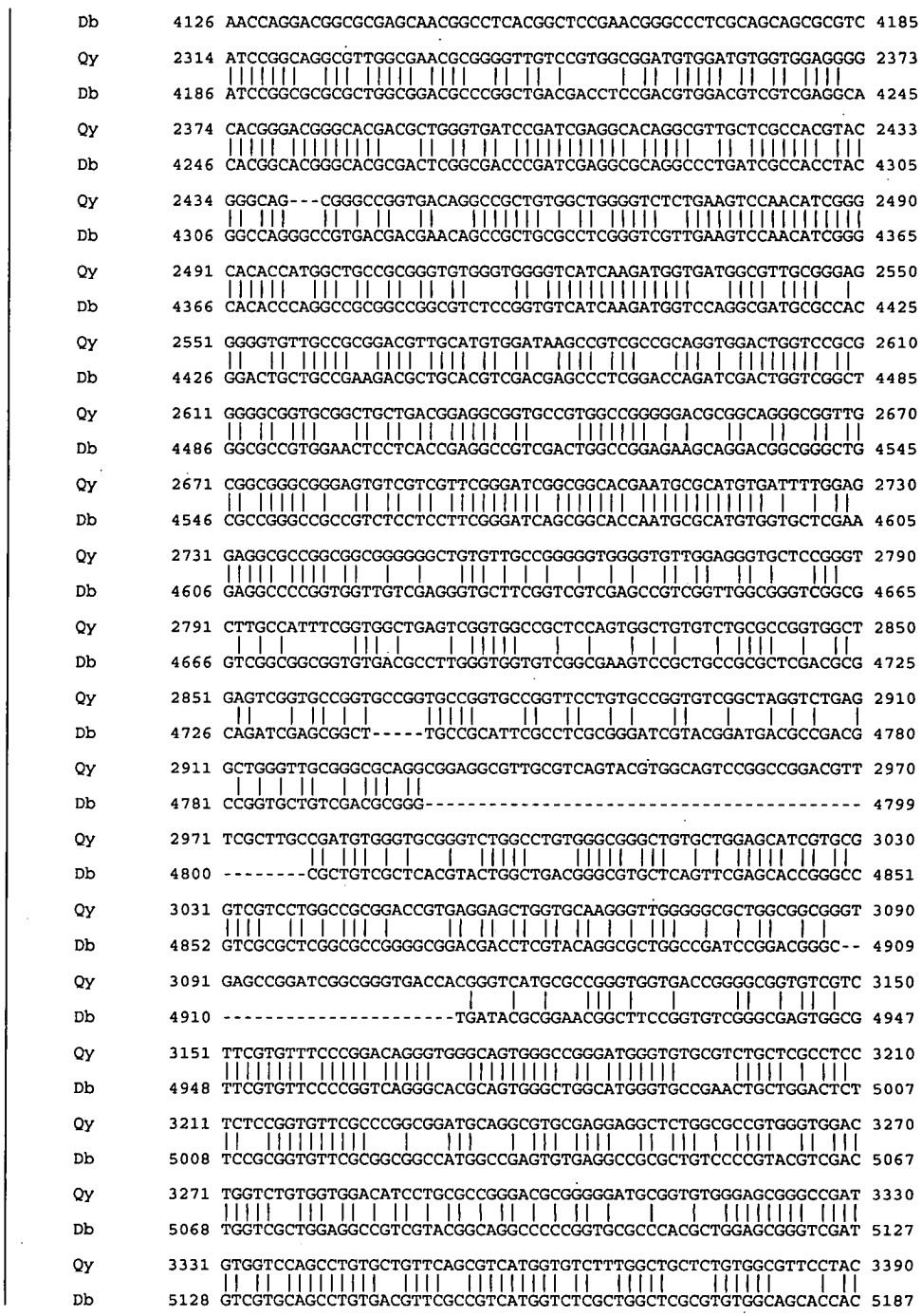





Db 7024 TCCGTGTCGAGCACTCTGGGCATCCCCGGTCAGGGCAACTACGCCCCGCACAACGCCTAC 7083
Qy 5608 CTCGACGCCCTCGCCCACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGG 5667
Db 7084 CTCGACGCCCTCGCGGCTCGCCGCCGGGCCACCGGCCGGTCCGCCGTCTCGGTGGCCTGG 7143
Qy 5668 GGCACCTGGCAAGGAAACGGACTCGCTGATTCGGACAAGGCCCGCGCATATCTCGACCGC 5727
Db 7144 GGACCGTGGGACGGTGGCGGCATGGCCGCCGGTGACGGCGTGGCCGAGCGGCTGCGCAAC 7203
Qy 5728 CGCGGGTTTCGACCCATGTCACCCGAGTTGGCCACGGCAGC---------- 5768
Db 7204 CACGGCGTGCCCGGCATGGACCCGGAACTCGCCCTGGCCGCACTGGAGTCCGCGCTCGGC 7263
Qy 5769 --------GGTCACGCAGGCGATCGCGGACACCGAACGGCCGTATGTCGTCATCGCCGAC 5820
Db 7264 CGGGACGAGACCGCGATCACCGTCGCGGACATCGACTGGGACCGCTTCTACCTCGCGTAC 7323
Qy 5821 ATCGACTGGAGCAAGATCGAACACACCTCTCAGACCAGCGACCTG------GTGAGCGCG 5874
Db 7324 TCCTCCGGTCGCCCGCAGCCCCTCGTCGAGGAGCTGCCCGAGGTGCGGCGCATCATCGAC 7383
Qy 5875 GCCCGGGAAAGGGAGCCAGCTGTCCAGCGCCCCACTCCACCGGCGG---------AGTTG 5925
Db 7384 GCACGGGACAGCGCCACGTCCGGACAGGGCGGGAGCTCCGCCCAGGGCGCCAACCCCCTG 7443
Qy 5926 CACAAAACGCTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTA 5985
Db 7444 GCCGAGCGGCTGGCCGCCGCGGCTCCCGGCGAGCGTACGGAGATCCTCCTCGGTCTCGTA 7503
Qy 5986 CGAGACCATGTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAG 6045
Db 7504 CGGGCGCAGGCCGCCGCCGTGCTCCGGATGCGTTCGCCGGAGGACGTCGCCGCCGACCGC 7563
Qy 6046 TCGTTCCGTGCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATC 6105
Db 7564 GCCTTCAAGGACATCGGCTTCGACTCGCTCGCCGGTGTCGAGCTGCGCAACAGGCTGACC 7623
Qy 6106 AAGGCAACAGGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAA 6165
Db 7624 CGGGCGACCGGGCTCCAGCTGCCCGCGACGCTCGTCTTCGACCACCCGACGCCGCTGGCC 7683
Qy 6166 CTCGCCGTACACCTGCAGAACCAACTGCGGGGCACAGCAGCGGAGTCG---------- 6213
Db 7684 CTCGTGTCGCTGCTCCGCAGCGAGTTCCTCGGTGACGAGGAGACGGCGGACGCCCGGCGG 7743
Qy 6214 ----------GCTCCTTCAGCGGCAGCCGTTACCGCCGAGGCTTCT 6249
Db 7744 TCCGCGGCGCTGCCCGCGACTGTCGGTGCCGGTGCCGGCGCCGGCGCCGGCACCGATGCC 7803
Qy 6250 GTCACCGAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCG 6309
Db 7804 GACGACGATCCGATCGCGATCGTCGCGATGAGCTGCCGCTACCCCGGTGACATCCGCAGC 7863
Qy 6310 GCGGACGACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACC 6369
Db 7864 CCGGAGGACCTGTGGCGGATGCTGTCCGAGGGCGGCGAGGGCATCACGCCGTTCCCCACC 7923
Qy 6370 GACCGCGGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGC 6429
Db 7924 GACCGCGGCTGGGACCTCGACGGCCTGTACGACGCCGACCCGGACGCGCTCGGCAGGGCG 7983
Qy 6430 TACACCCGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGC 6489
Db 7984 TACGTCCGCGAGGGCGGGTTCCTGCACGACGCGGCCGAGTTCGACGCGGAGTTCTTCGGC 8043
Qy 6490 ATCAGCCCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCC 6549
Db 8044 GTCTCGCCGCGCGAGGCGCTGGCCATGGACCCGCAGCAGCGGATGCTCCTGACGACGTCC 8103
Qy 6550 TGGGAAACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGA 6609
Db 8104 TGGGAGGCCTTCGAGCGGGCCGGCATCGAGCCGGCATCGCTGCGCGGCAGCAGCACCGGT 8163
Qy 6610 GTCTTCACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCA 6669
Db 8164 GTCTTCATCGGCCTCTCCTACCAGGACTACGCGGCCCGCGTCCCGAACGCCCCGCGTGGC 8223
Qy 6670 ACCGATGGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTAC 6729
Db 8224 GTGGAGGGTTACCTGCTGACCGGCAGCACGCCGAGCGTCGCGTCGGGCCGTATCGCGTAC 8283
Qy 6730 ACGTTTGGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTG 6789
Db 8284 ACCTTCGGTCTCGAAGGGCCCGCGACGACCGTCGACACCGCCTGCTCGTCGTCGCTGACC 8343
Qy 6790 GCTTTGCATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGG 6849
Db 8344 GCCCTGCACCTGGCGGTGCGGGCGCTGCGCAGCGGCGAGTGCACGATGGCGCTCGCCGGT 8403
Qy 6850 GGTGTGACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTG 6909
Db 8404 GGCGTGGCGATGATGGCGACCCCGCACATGTTCGTGGAGTTCAGCCGTCAGCGGGCGCTC 8463
Qy 6910 GCCGCGGACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAG 6969
Db 8464 GCCCCGGACGGCCGCAGCAAGGCCTTCTCGGCGGACGCCGACGGGTTCGGCGCCGCGGAG 8523
Qy 6970 GGTGTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTC 7029
Db 8524 GGCGTCGGCCTGCTGCTCGTGGAGCGGCTCTCGGACGCGCGGCGCAACGGTCACCCGGTG 8583
Qy 7030 CTGGCCGTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCG 7089
Db 8584 CTCGCCGTGGTCCGCGGTACCGCCGTCAACCAGGACGGCGCCAGCAACGGGCTGACCGCG 8643
Qy 7090 CCCAACGGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCG 7149
Db 8644 CCCAACGGACCCTCGCAGCAGCGGGTGATCCGGCAGGCGCTCGCCGACGCCCGGCTGGCA 8703
Qy 7150 GCCGGTGATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATC 7209
Db 8704 CCCGGCGACATCGACGCCGTCGAGACGCACGGCACGGGAACCTCGCTGGGCGACCCCATC 8763
Qy 7210 GAGGCCCAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGG 7269
Db 8764 GAGGCCCAGGGCCTCCAGGCCACGTACGGCAAGGAGCGGCCCGCGGAACGGCCGCTCGCC 8823
Qy 7270 CTGGGCTCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTG 7329
Db 8824 ATCGGCTCCGTGAAGTCCAACATCGGACACACCCAGGCCGCGGCCGGTGCGGCGGGCATC 8883
Qy 7330 ATCAAGATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAG 7389
Db 8884 ATCAAGATGGTCCTCGCGATGCGCCACGGCACCCTGCCGAAGACCCTCCACGCCGACGAG 8943
Qy 7390 CCGTCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCC 7449
Db 8944 CCGAGCCCGCACGTCGACTGGGCGAACAGCGGCCTGGCCCTCGTCACCGAGCCGATCGAC 9003
Qy 7450 TGGCCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGC 7509
Db 9004 TGGCCGGCCGGCACCGGTCCG---CGCCGCGCCGCCGTCTCCTCCTTCGGCATCAGCGGG 9060
Qy 7510 ACCAACGCCCACGTCATCCTCGAAGAAGCACCCGCCGACGACGTT---------- 7554
Db 9061 ACGAACGCGCACGTCGTGCTGGAGCAGGCGCCGGATGCTGCTGGTGAGGTGCTTGGGGCC 9120
Qy 7555 ----------CCGGGGGGACCACCC 7569
Db 9121 GATGAGGTGCCTGAGGTGTCTGAGACGGTAGCGATGGCTGGGACGGCTGGGACCTCCGAG 9180
Qy 7570 GCCGGCGAGGGTGACGCGGGCAGCGACGATGAGGCTGCTGCCGGCAGTCCTGGGGTGT-- 7627
Db 9181 GTCGCTGAGGGCTCTGAGGCCTCCGAGGCCCCCGCGGCCCCCGGCAGCCGTGAGGCGTCC 9240


Qy 3391 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 3450
Db 5188 GGTGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 5247
Qy 3451 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 3510
Db 5248 GCCGGAGCCCTGCCCCTGGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 5307
Qy 3511 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 3567
Db 5308 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAACGAGGACGCCGTC 5367
Qy 3568 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 3627
Db 5368 CTGGAGCGACTGAGTGA------CTTCGACGGGCTGTCCGTCGCCGCCGTCAACGGGCCC 5421
Qy 3628 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 3687
Db 5422 ACCGCCACTGTCGTGTCGGGTGACCCCGTACAGATCGAAGAGCTTGCTCAGGCGTGCAAG 5481
Qy 3688 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 3747
Db 5482 GCGGACGGATTCCGCGCGCGGATCATTCCCGTCGACTACGCGTCCCACAGCCGGCAGGTC 5541
Qy 3748 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 3807
Db 5542 GAGATCATCGAGAGCGAGCTCGCCCAGGTCCTCGCCGGTCTCAGCCCGCAGGCCCCGCGC 5601
Qy 3808 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 3867
Db 5602 GTGCCGTTCTTCTCGACGCTCGAAGGCACCTGGATCACCGAGCCCGTCCTCGACGGCACC 5661
Qy 3868 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 3927
Db 5662 TACTGGTACCGCAACCTCCGTCACCGCGTCGGCTTCGCCCCCGCCATCGAGACCCTGGC- 5720
Qy 3928 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 3987
Db 5721 ----------CGTCGACGAGGGCTTCACGCACTTCGTCGAGGTCAGCGCCCAC 5763
Qy 3988 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 4047
Db 5764 CCCGTCCTCACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGCACCCTCCGTCGCGAA 5823
Qy 4048 GACAACGACACCCGCCGCTTCCTCACCGCCCTCGCCCACACCCATACCACCGGCATCGGC 4107
Db 5824 CAGGGAGGCCAAGAGCGTCTGGTCACCTCGCTCGCCGAGGCGTGGGTCAACGGGCTTCCC 5883
Qy 4108 ACACCCACCACCTGGCACCACCACTACACCCACCACCACACCCACCCCCACCCCCACACG 4167
Db 5884 GTGGCATGGACTTCGC----------TCCTGCCCGCCACGGCCTCC 5919
Qy 4168 CACCTCGACCTGCCCACCTACCCCTTCCAACACCAGCACTACTGGCTCGAGAGCTCACAG 4227
Db 5920 CGCCCCGGTCTGCCCACCTACGCCTTCCAGGCCGAGCGCTACTGGCTCGAGAACACTCCC 5979
Qy 4228 CCGGGTGCCGGATCCGGTTCGGGTGCCGGTGCCGGTTCGGGTGCCGGTTCCGGGCGGGCA 4287
Db 5980 GCCG----------CCCTGGCCACCG 5995
Qy 4288 GGGACTGCGGGCGGGACGGCAGAGGTGGAGTCGCGGTTCTGGGACGCGGTGGCCCGCCAG 4347
Db 5996 GCGACGACTGGCGCTACCGCATCGACTG-GAAGCGCCTCCCGGCCGCCGAGGGGTCCGAG 6054
Qy 4348 GACCTGGAAACGGTCGCGACCACACTCGCCGTGCCCCCCTCCGCCGGCCTGGACACGGTG 4407
Db 6055 CGCACCGGCCTGTCCGGCCGCTGGCTCGCCGTCACGCCGGAGGACCACTCCGCGCAGGCC 6114
Qy 4408 GTGCCCGCACTCTCCGCCTGGCACCGCCACCAACACGACCAAGCCCGCATCAACACCTGG 4467
Db 6115 GCCGCCGTGCTCACCGCGCTGGTCGACGCCGGGGCGAAGGTCGAGGTGCTGACGGCCGGG 6174
Qy 4468 ACCTACCAGGAAACCTGGAAACCCCTCACCCTCCCCACCACCCACCAACCCCACCAAACC 4527
Db 6175 GCGGACGACGACCGTGAGGCCCTCGCCGCCCGGCTCACCGCACTGACG---------- 6222
Qy 4528 TGGCTCATCGCCATCCCCGAAACCCAGACCCACCACCCCCACATCACCAACATCCTCACC 4587
Db 6223 ---------- 6222
Qy 4588 AACCTCCACCACCACGGCATCACCCCCATCCCCCTCACCCTCAACCACACCCACACCAAC 4647
Db 6223 ---------- 6222
Qy 4648 CCCCAACACCTCCACCACACCCTCCACCACACCCGACAACAAGCCCAAAACCACACCACC 4707
Db 6223 ----------ACCGGT 6228
Qy 4708 GGAGCCATCACCGGCCTGCTCTCCCTCCTCGCCCTCGACGAAACACCCCACCCCCACCAC 4767
Db 6229 GACGGCTTCACCGGCGTGGTCTCGCTCCTCG---------- 6259
Qy 4768 CCCCACACACCCACCGGCACCCTCCTCAACCTCACCCTCACCCAAACCCACACCCAAACC 4827
Db 6260 ----------ACGGACTCGTACCGCAGGTCGCCTGGGTCCAGGCGCTCGGCGACGCC 6306
Qy 4828 CACCCACCAACCCCCCTCTGGTACGCCACCACCAACGCCACCACCACCCACCCCAACGAC 4887
Db 6307 GGAATCAAGGCGCCCCTGTGGTCCGTCACCCAGGGCGCGGTCTCCGTCGGACGTCTCGAC 6366
Qy 4888 CCCCTCACACACCCCACCCAAGCCCAAACCTGGGGACTCGCCCGCACCACCCTCCTCGAA 4947
Db 6367 ACCCCCGCCGACCCCGACCGGGCCATGCTCTGGGGCCTCGGCCGCGTCGTCGCCCTTGAG 6426
Qy 4948 CACCCCACCCACACCGCCGGAATCATCGACCTCCCCACCACCCCCACCCCCCACACCCTC 5007
Db 6427 CACCCCGAACGCTGGGCCGGCCTCGTCGACCTCCCCGCCCAGCCCGATGCCGCCGCCCTC 6486
Qy 5008 CAGCACCTCACCCAAACCCTCACCCAACCCCACCACCAAACCCAACTCGCCATCCGCACC 5067
Db 6487 GCCCACCTCGTCACCGCACTCTCCGGCGCCACCGGCGAGGACCAGATCGCCATCCGCACC 6546
Qy 5068 ACCGGCACCCACACCCGCCGCCTCACCCCCACCACCCTCACCCCCACACACCAACCACCC 5127
Db 6547 ACCGGACTCCACGCCCGCCGCCTCGCCCGCGCACCCCTCCACGGACGTCGGCCCACCCGC 6606
Qy 5128 ACCCCCACCCCCCACGGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCAC 5187
Db 6607 GACTGGCAGCCCCACGGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCAC 6666
Qy 5188 CTCACCCACCACCTCACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACC 5247
Db 6667 ---GCCGCACGCTGGATGGCCCACCACGGAGCCGAACACCTCCTCCTCGTCAGCCGCAGC 6723
Qy 5248 GGCCCCCACACCCCCCACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCAC 5307
Db 6724 GGCGAACAAGCCCCCGGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGC 6783
Qy 5308 CTCACCATCACCACCTGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACC 5367
Db 6784 GTCACCATCGCCGCCTGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCC 6843
Qy 5368 ATCCCCCCACAACACCCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCC 5427
Db 6844 ATCCCCGCCGAGACGCCCCTCACCGCCGTCGTCCACACCGCCGGCGCGCTCGACGACGGC 6903
Qy 5428 ACCCTCACCAACCTCACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGC 5487
Db 6904 ATCGTGGACACGCTGACCGCCGAGCAGGTCCGGCGGGCCCACCGTGCGAAGGCCGTCGGC 6963
Qy 5488 GCCCACCTCCTCCACCAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCC 5547
Db 6964 GCCTCGGTGCTCGACGAGCTGACCCGGGACCTCGACCTCGACGCGTTCGTGCTCTTCTCG 7023
Qy 5548 TCCGCCGCCGCCACCTTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTAC 5607





```
Qy      14361 CGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAAC 14420
              ||||| ||||  | |||||  | |  |||||   | | |||||  | ||  |  |
Db      19143 CGCCTCCCACAGCGCCCACGTCGAGACCATCGAGAGCGAACTCGCCGAGGTCCTCGCCGG 19202

Qy      14421 CCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACT 14480
               ||||  |    |  ||||  |    ||  || || |  |   |        ||
Db      19203 GCTCAGCCCGCGGACACCTGAGGTGCCGTTCTTCTCGACACTCGAAGGCGCCTGGATCAC 19262

Qy      14481 CCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCAC 14540
              |    |    | | || | ||| |       |||||||  |  | ||  || |  |  | |
Db      19263 CGAGCCGGTGCTCGACGGCACCTACTGGTACCGCAACCTCCGCCACCGCGTCGGCTTCGC 19322

Qy      14541 CCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCT 14600
              ||   || || |    | || |  |  |     | ||   | |     |||   || |
Db      19323 CCCCGCCGTCGAGACCCTCGCCACCGACGAAGGCTTCACCCACTTCATCGAGGTCAGCGC 19382

Qy      14601 CACCACCCTCACCCACGACAACCTCCCCAACACCCCCACCACCACCCTCACCCTCACCCA 14660
              |  | || ||  |  |  | |||||| | |||  ||||  |  |  ||||||   ||
Db      19383 CCACCCCGTCCTCACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGCACCCT---CCG 19439

Qy      14661 CCCCCACCACCACCCCCAAACCCACCTCCTCACCAACCTCGCCAAAACCACCACCACCTG 14720
              || | | ||      |||    |  ||  |||||   |||||| || ||   |||| | |
Db      19440 CCGCGAACAGGGAGGCCAGGAGCGTCTGGTCACCTCACTCGCCGAAGCCTGGACCAACGG 19499

Qy      14721 GCACCCCCACCACTACACCCACCACCACAACCAACCCCACACCCACACCCACCTCGACCT 14780
               | | ||  | |||  | | |  ||    ||   ||   |||  |  |||| ||| ||
Db      19500 CCTCACCATCGACTGGGCGCCCGTCC---TCCCCACCGCAACCGGCCACCACCCCGAGCT 19556

Qy      14781 CCCCACCTACCCCTTCCAACACCACCACTACTGGCTCCA 14819
              |||||||||| ||||||| | ||  ||||||||||||||
Db      19557 CCCCACCTACGCCTTCCAGCGCCGTCACTACTGGCTCCA 19595
```

RESULT 6
US-09-860-846-30
; Sequence 30, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0 .
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-30

Query Match 7.6%; Score 2342; DB 9; Length 13842;
Best Local Similarity 54.3%; Pred. No. 0;
Matches 6606; Conservative 0; Mismatches 4530; Indels 1026; Gaps 52;

```
Qy          8 GGATGGACGGCGGGGAAGAACCCCGCCCTGCGGCAGGGGAGGTCCTCGGAGTGGCCGACG 67
              ||  |||||    ||  | | | |   | ||| | | ||   |  || | |  | |
Db       1820 GGCCGGACGATCTGGCGGCAGCGCTGGCCGCGCCTGAGGGTCTGGTCCGGGGCGTGGCTT 1879

Qy         68 AGGCGGACGGCGGCGTCGTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGG 127
                |  | |||  | || |  ||||| || ||||||||||||| |||| ||| | |||||||
Db       1880 CCGGTGTCGGGCGAGTGGCGTTCGTGTTCCCCGGGCAGGGCACGCAGTGGGCCGGCATGG 1939

Qy        128 GAAGGGAACTTCTCGACGCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCG 187
              |    ||||| || ||| |||||  || |||  || |  | |    ||  |||||  ||||
Db       1940 GTGCCGAACTGCTGGACTCTTCCGCGGTGTTCGCGGCGGCCATGGCCGAATGCGAGGCCG 1999
```

```
Qy        188 CGTTCGCGCCCTACGTCGACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTC 247
              |  || | ||  |||||||||||||||| |||||   ||  | ||| |  | || |  || |
Db       2000 CACTCTCCCCGTACGTCGACTGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGC 2059

Qy        248 CCGGGCTGGACCGGGTGGACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGG 307
              ||  |||||| |||||  || |||||  |||||        |||||||||||||  |||| ||||
Db       2060 CCACGCTGGAGCGGGTCGATGTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGG 2119

Qy        308 CCGCCCTCTGGCGCTCGCAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCG 367
              |   | | ||||     || ||||||  |||    |||||  | || |||   | |||||
Db       2120 CTCGCGTGTGGCAGCACCACGGGGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCG 2179

Qy        368 AGATCGCGGCAGCCCACGTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGA 427
              |||||||  || ||  |||||  | || |  |||| |||||  |||||| ||  || ||||
Db       2180 AGATCGCCGCCGCGTACGTCGCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGA 2239

Qy        428 CGCTTTGGAGCCAGGCACAGACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCG 484
              |  ||  | ||| ||  |      || |  ||| |||||  |  ||||    |  | ||| |||
Db       2240 CCCTGCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCG 2299

Qy        485 CCGCCACGCCGGATGAGCTCCTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGC 544
              |   |   ||| |  |||||   ||| | ||   ||    ||
Db       2300 CGCTGAGCGAGGACGCCGTCCTGGAGCGACTGGCCGGGTTCGACG--------------- 2344

Qy        545 GGCTCGCCGTCGCAGCCGTCAACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGG 604
              ||||  |||||||| || ||  |||||  |||   ||| || || |||||||  || ||
Db       2345 GGCTGTCCGTCGCCGCTGTGAACGGGCCCACCGCCACCGTGGTCTCCGGTGACCCCGTAC 2404

Qy        605 CCGTCGCGGACCTGGTGGCCGACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGG 664
               |||  || || |    |      ||||    ||| ||  ||||  | || || |
Db       2405 AGATCGAAGAGCTTGCTCGGGCGTGTGAGGCCGATGGGGTCCGTGCGCGGGTCATTCCCG 2464

Qy        665 TGGACGTTCCCGCCCACTCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCC 724
              | |||    |  |||||  ||  | | |  |  |||||||     | |||  |  | ||
Db       2465 TCGACTACGCGTCCCACAGCCGGCAGGTCGAGATCATCGAGAGCGAGCTCGCCGAGGTCC 2524

Qy        725 TGCTGCCCATCACCCCACGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCC 784
              |          ||| ||| |   |  | ||| | || |||  |||  |  |    |||| |
Db       2525 TCGCCGGGCTCAGCCCGCAGGCTCCGCGCGTGCCGTTCTTCTCGACACTCGAAGGCGCCT 2584

Qy        785 GCCTCGACACCCGCGAGCTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCC 844
              |  ||  |   | || ||| ||||  | ||||||||||||||| ||          ||
Db       2585 GGATCACCGAGCCCGTGCTCGACGGCGGCTACTGGTACCGCAACCTGCGCCATCGTGTGG 2644

Qy        845 GGTTCGAGCCCGCCGCCCGG---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGA 901
              | ||||  || ||||  |  |  |||     |  ||||   ||     |||||||||
Db       2645 GCTTCGCCCCGGCCGTCGAGACCCTGGCCACCGACGAGGGCTTCACCCACTTCGTCGAGG 2704

Qy        902 TGAGCCCGCACCCGGTGCTGACCATGGGCCTCCAGGAGCTCGCCCCGGACCTGGG----- 956
              | ||| | ||||| || || |||||| ||||  | |  || | | |  ||||
Db       2705 TCAGCGCCCACCCCGTCCTCACCATGGCCCTCCCCGGGACCGTCACCGGTCTGGCGACCC 2764

Qy        957 ----------------------------------------------CGACACCACCGGCA 970
                                                           ||| |   | ||
Db       2765 TGCGTCGCGACAACGGCGGTCAGGACCGCCTAGTCGCCTCCCTCGCCGAAGCATGGGCCA 2824

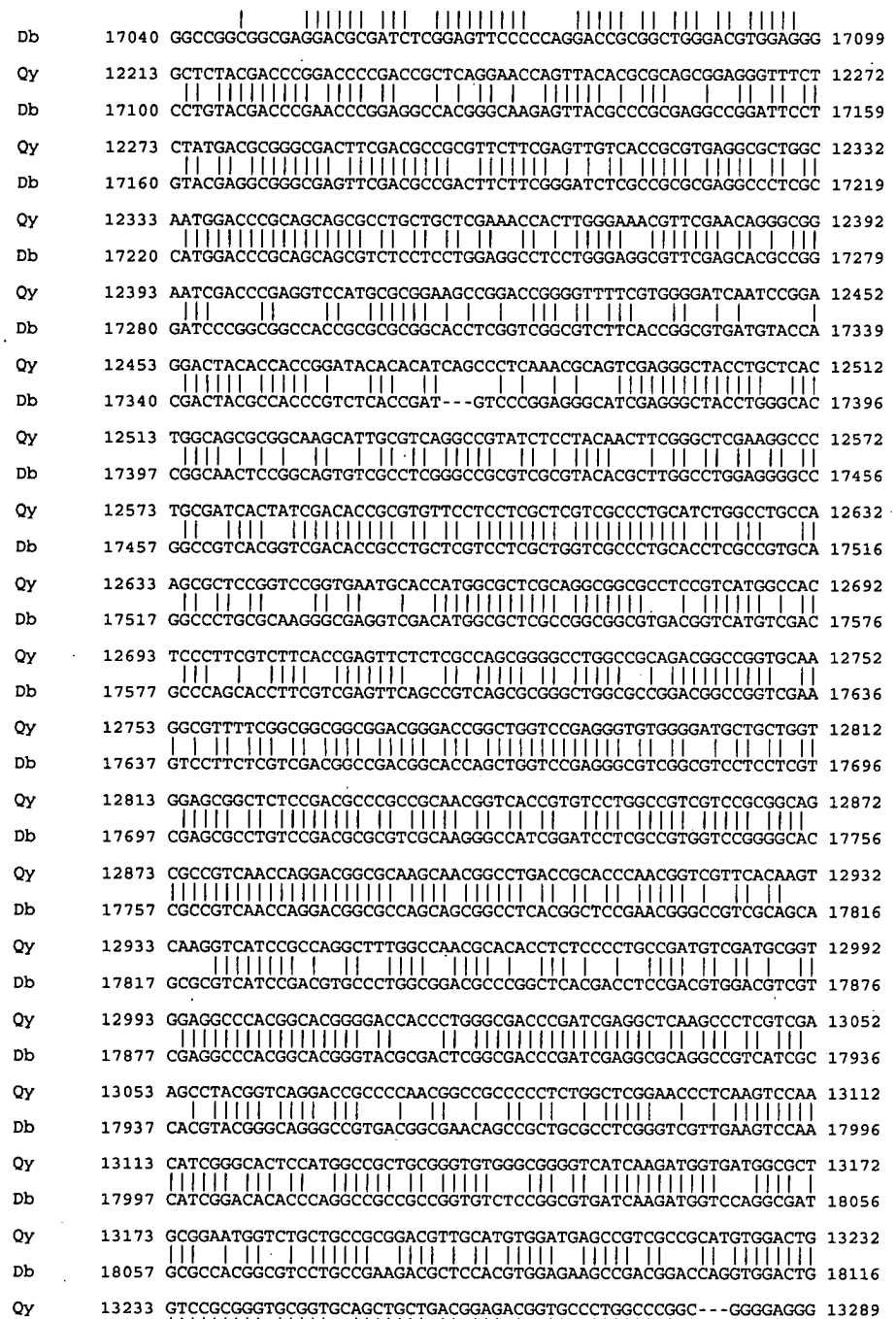
Qy        971 CCGCCGACACCGTGATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCC 1030
               ||   | | ||   | | || |||| | || |  |  |  |  ||  ||  | ||
Db       2825 ACGGACTCGCGGTCGACTGGAGCCCGCTCCTCCCCTCCGCGACCGGCCACCACTCCGACC 2884

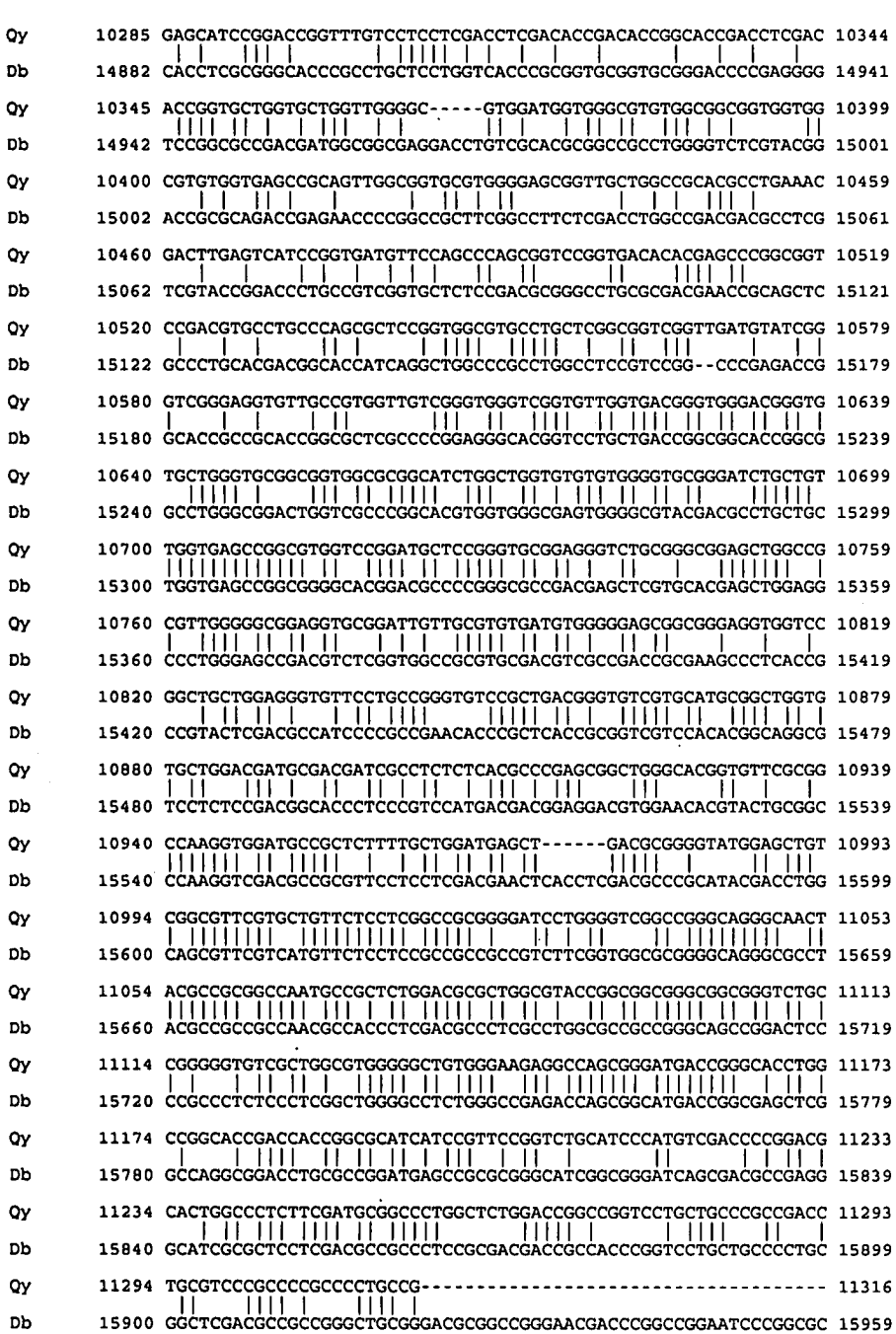
Qy        1031 TGACGTCTCTCGC---------------CCAACTACGGGGGCATGGTGAGACGTCGGCGA 1075
               |  |  |  |||                |||  | |||  |     ||| ||  ||
Db       2885 TCCCCACCTACGCGTTCCAGACCGAGCGCCACTGGCTGGGCGAGATCGAGGCGCTCGCCC 2944

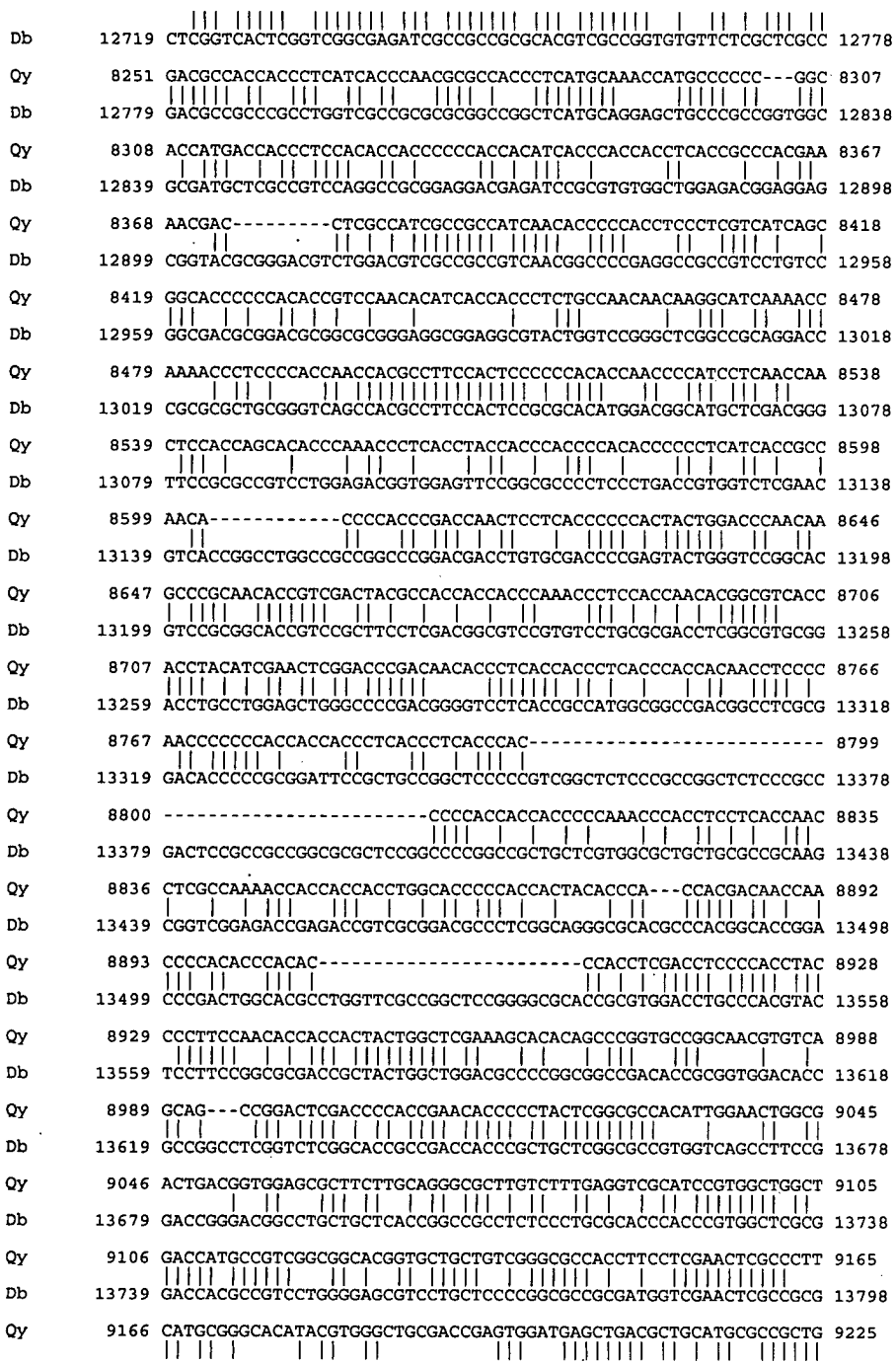
Qy        1076 CCACCGTCCTCTCGGCACGCCTGACCGCGCTGTCCCCCACGCAGCAGCAGTCGCTGCTCC 1135
               | | |  |||    | | |||  |   || |||| ||  |     || |
Db       2945 CGGCGGGCGAGCCGGCGGTGCAGCCCGCCGTCCTCCGCACGGAGGCGGCCGAGCCGGCGG 3004

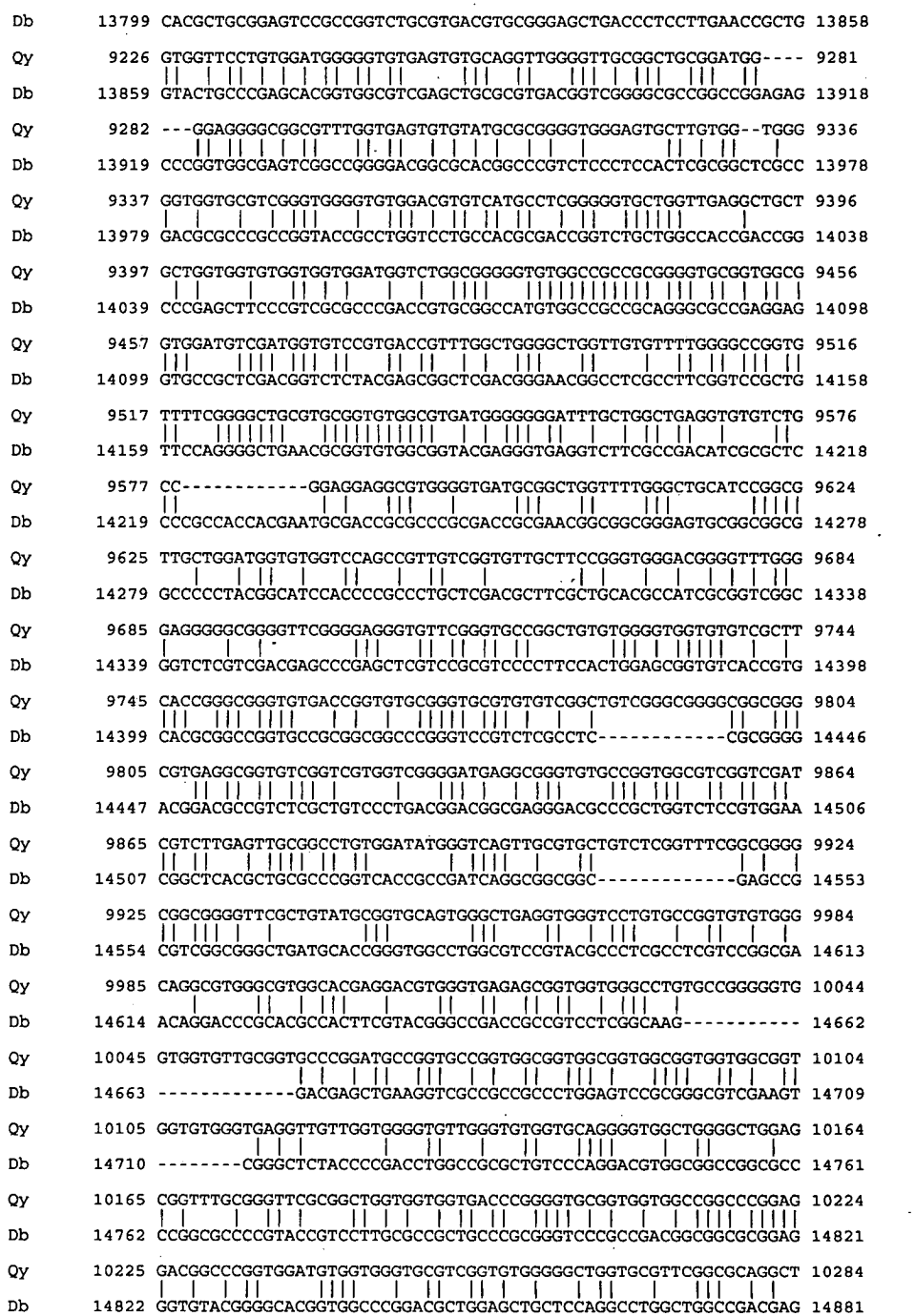
Qy        1136 TGGACCTGGTGCGCGCCCACACCATGGCGGTGCTGAACGACGACGGAAACGAGCGCACCG 1195
               |  |     |  ||  ||      | | | ||| || | | | |     ||    ||
Db       3005 AGCTCGACCGGGACGAGCAGCTGCGCGTGATCCTGGACAAGGTCCGGGCGCAGACGGCCC 3064

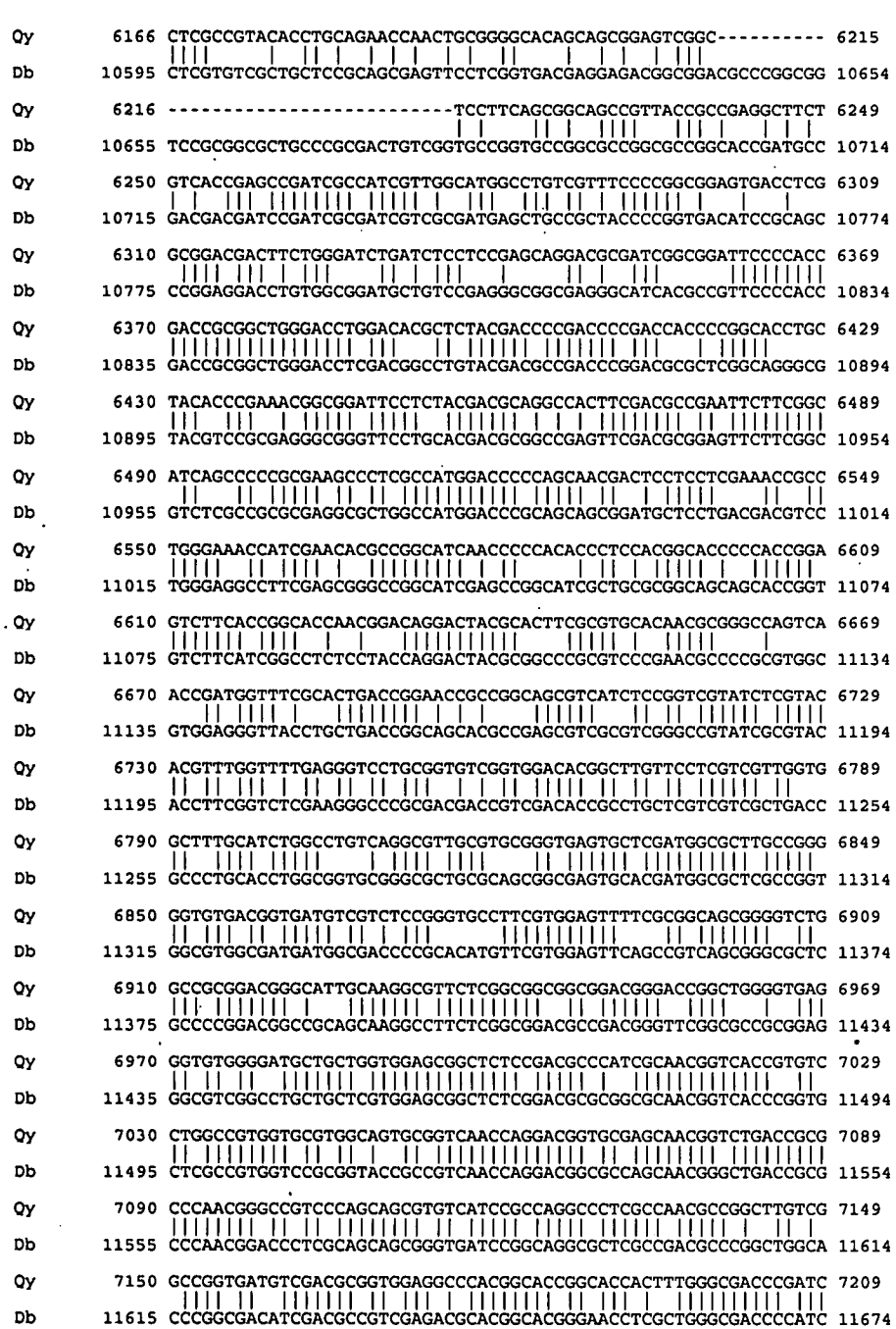
Qy        1196 CGTCGGATGCCGGCCCATCGG--------------------CGAGTTTCGCCCACCTCG 1234
```

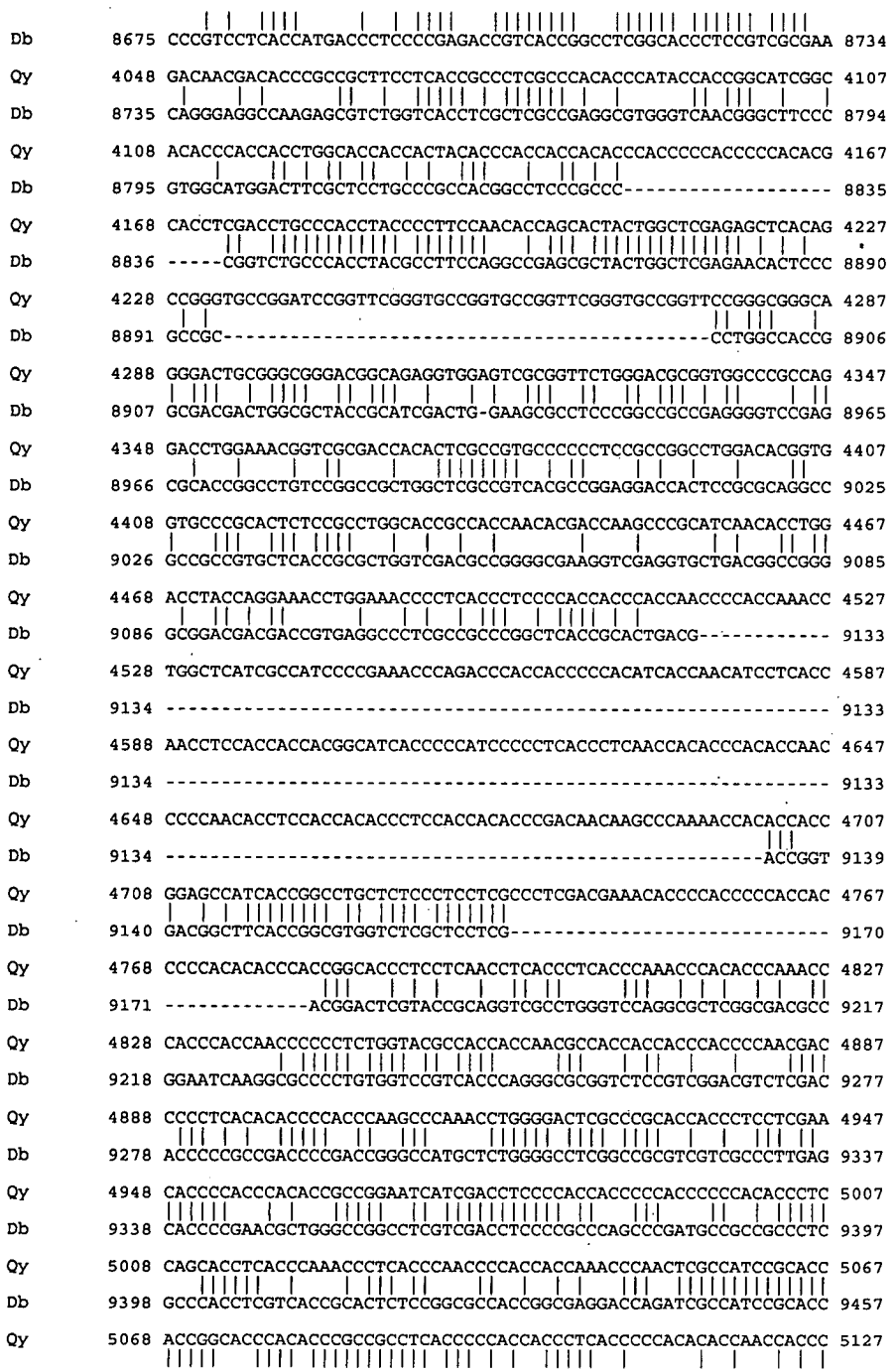

Db 17040 GGCCGGCGGCGAGGACGCGATCTCGGAGTTCCCCCAGGACCGCGGCTGGGACGTGGAGGG 17099
Qy 12213 GCTCTACGACCCGGACCCCGACCGCTCAGGAACCAGTTACACGCGCAGCGGAGGGTTTCT 12272
Db 17100 CCTGTACGACCCGAACCCGGAGGCCACGGGCAAGAGTTACGCCCGCGAGGCCGGATTCCT 17159
Qy 12273 CTATGACGCGGGCGACTTCGACGCCGCGTTCTTCGAGTTGTCACCGCGTGAGGCGCTGGC 12332
Db 17160 GTACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCGCCGCGCGAGGCCCTCGC 17219
Qy 12333 AATGGACCCGCAGCAGCGCCTGCTGCTCGAAACCACTTGGGAAACGTTCGAACAGGGCGG 12392
Db 17220 CATGGACCCGCAGCAGCGTCTCCTCCTGGAGGCCTCCTGGGAGGCGTTCGAGCACGCCGG 17279
Qy 12393 AATCGACCCGAGGTCCATGCGCGGAAGCCGGACCGGGGTTTTCGTGGGGATCAATCCGGA 12452
Db 17280 GATCCCGGCGGCCACCGCGCGCGGCACCTCGGTCGGCGTCTTCACCGGCGTGATGTACCA 17339
Qy 12453 GGACTACACCACCGGATACACACATCAGCCCTCAAACGCAGTCGAGGGCTACCTGCTCAC 12512
Db 17340 CGACTACGCCACCCGTCTCACCGAT---GTCCCGGAGGGCATCGAGGGCTACCTGGGCAC 17396
Qy 12513 TGGCAGCGCGGCAAGCATTGCGTCAGGCCGTATCTCCTACAACTTCGGGCTCGAAGGCCC 12572
Db 17397 CGGCAACTCCGGCAGTGTCGCCTCGGGCCGCGTCGCGTACACGCTTGGCCTGGAGGGGCC 17456
Qy 12573 TGCGATCACTATCGACACCGCGTGTTCCTCCTCGCTCGTCGCCCTGCATCTGGCCTGCCA 12632
Db 17457 GGCCGTCACGGTCGACACCGCCTGCTCGTCCTCGCTGGTCGCCCTGCACCTCGCCGTGCA 17516
Qy 12633 AGCGCTCCGGTCCGGTGAATGCACCATGGCGCTCGCAGGCGGCGCCTCCGTCATGGCCAC 12692
Db 17517 GGCCCTGCGCAAGGGCGAGGTCGACATGGCGCTCGCCGGCGGCGTGACGGTCATGTCGAC 17576
Qy 12693 TCCCTTCGTCTTCACCGAGTTCTCTCGCCAGCGGGGCCTGGCCGCAGACGGCCGGTGCAA 12752
Db 17577 GCCCAGCACCTTCGTCGAGTTCAGCCGTCAGCGCGGGCTGGCGCCGGACGGCCGGTCGAA 17636
Qy 12753 GGCGTTTTCGGCGGCGGCGGACGGGACCGGCTGGTCCGAGGGTGTGGGGATGCTGCTGGT 12812
Db 17637 GTCCTTCTCGTCGACGGCCGACGGCACCAGCTGGTCCGAGGGCGTCGGCGTCCTCCTCGT 17696
Qy 12813 GGAGCGGCTCTCCGACGCCCGCCGCAACGGTCACCGTGTCCTGGCCGTCGTCCGCGGCAG 12872
Db 17697 CGAGCGCCTGTCCGACGCGCGTCGCAAGGGCCATCGGATCCTCGCCGTGGTCCGGGGCAC 17756
Qy 12873 CGCCGTCAACCAGGACGGCGCAAGCAACGGCCTGACCGCACCCAACGGTCGTTCACAAGT 12932
Db 17757 CGCCGTCAACCAGGACGGCGCCAGCAGCGGCCTCACGGCTCCGAACGGGCCGTCGCAGCA 17816
Qy 12933 CAAGGTCATCCGCCAGGCTTTGGCCAACGCACACCTCTCCCCTGCCGATGTCGATGCGGT 12992
Db 17817 GCGCGTCATCCGACGTGCCCTGGCGGACGCCCGGCTCACGACCTCCGACGTGGACGTCGT 17876
Qy 12993 GGAGGCCCACGGCACGGGGACCACCCTGGGCGACCCGATCGAGGCTCAAGCCCTCGTCGA 13052
Db 17877 CGAGGCCCACGGCACGGGTACGCGACTCGGCGACCCGATCGAGGCGCAGGCCGTCATCGC 17936
Qy 13053 AGCCTACGGTCAGGACCGCCCCAACGGCCGCCCCCTCTGGCTCGGAACCCTCAAGTCCAA 13112
Db 17937 CACGTACGGGCAGGGCCGTGACGGCGAACAGCCGCTGCGCCTCGGGTCGTTGAAGTCCAA 17996
Qy 13113 CATCGGGCACTCCATGGCCGCTGCGGGTGTGGGCGGGGTCATCAAGATGGTGATGGCGCT 13172
Db 17997 CATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCGTGATCAAGATGGTCCAGGCGAT 18056
Qy 13173 GCGGAATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCGCCGCATGTGGACTG 13232
Db 18057 GCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGAGAAGCCGACGGACCAGGTGGACTG 18116
Qy 13233 GTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCCGGC---GGGGAGGG 13289
Db 18117 GTCCGCGGGCGCGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCGGACAAGGGCGACGG 18176
Qy 13290 GCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAACGCCCACGTCAT 13349
Db 18177 CGGACTGCGCAGGGCCGCGGTCTCCTCCTTCGGCGTCAGCGGGACGAACGCGCACGTCGT 18236
Qy 13350 CCTCGAGGAAGCACCCGCCCACAACATCCCGTCAGACACACCCGCCGACGACGTCCCGGG 13409
Db 18237 GCTCGAAGAGGCCCCGG----------CGGCCGAGG 18262
Qy 13410 AGAATCAGCCGCCGACGAGGATGCCGGTAGTGGCGATGAGGCTGCTGCCGGCAGTCCAGG 13469
Db 18263 AGACCCCTGCCTCCGAGGCGACCCCGGCCGTCGAGCCGT--------CGGTCGGCGCCGG 18314
Qy 13470 GGTGTGGCCGTGGCTGGTGTCGGCCAAGTCGCAGCCGGCCCTGCGCGCCCAGGCCCAGGC 13529
Db 18315 CCTGGTGCCGTGGCTGGTGTCGGCGAAGACTCCGGCCGCGCTGGACGCCCAGATCGGACG 18374
Qy 13530 CCTGCACGCCCACCTCACCGACCACC------CCGGCCTCGACCTCGCCGACGTCGGGTA 13583
Db 18375 CCTCGCCGCGTTCGCCTCGCAGGGCCGTACGGACGCCGCCGATCCGGGCGCGGTCGCTCG 18434
Qy 13584 CACCCTCGCCCACGCCCGCGCCGTGTTCGACCACCGCGCCACCCTCATCGCCGCCGACCG 13643
Db 18435 CGTACTGGCCGGCGGGCGCGCCGAGTTCGAGCACCGGGCCGTCGTGCTCGGCACCGGACA 18494
Qy 13644 CGACACCTTCCTGCAAGCACTCCAGGCACTCGCCGCAGGCGAACCCCACCCCGCCGTCAT 13703
Db 18495 GGACGATTTCGCGCAGGCGCTGACCGCTCCGGAAGGACTGATACGCGGCACGCCCTC--- 18551
Qy 13704 CCACAGCAGCGCCCCAGGCGGGACCGGGACCGGGGAGGCCGCAGGAAAGACCGCATTCAT 13763
Db 18552 ----------GGACGTGGGCCGGGTGGCGTTCGT 18575
Qy 13764 CTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCACACCCACCC 13823
Db 18576 GTTCCCCGGTCAGGGCACGCAGTGGGCCGGGATGGGCGCCGAACTCCTCGACGTGTCGAA 18635
Qy 13824 CGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTCGACCACCC 13883
Db 18636 GGAGTTCGCGGCGGCCATGGCCGAGTGCGAGAGCGCGCTCTCCCGCTATGTCGACTGGTC 18695
Qy 13884 CCTCCTCCCCCTCCTCACCCAGGACCCCAACACCCAGGACACCACCACCCTCGAAGAAGC 13943
Db 18696 GCTGGAGGCCGTCGTCCGGCAGGCGCC----------GGG 18725
Qy 13944 GGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGC 14003
Db 18726 CGCGCCCACGCTGGAGCGGGTCGACGTCGTCCAGCCCGTGACCTTCGCTGTCATGGTTTC 18785
Qy 14004 CCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTC 14063
Db 18786 GCT---GGCGAAGGTCTGGCAGCACCACGGCGTGACGCCGCAGGCCGTCGTCGGCCACTC 18842
Qy 14064 CCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCAC 14123
Db 18843 GCAGGGCGAGATCGCCGCCGCGTACGTCGCCGGTGCCCTCACCCTCGACGACGCCGCCCG 18902
Qy 14124 CCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCAC---CATGACCAC 14180
Db 18903 CGTCGTCACCCTGCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAAGGGCGGCATGAT 18962
Qy 14181 CCTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGC 14240
Db 18963 CTCCCTCGCCCTCAGCGAGGAAGCCACCCGGCAGCGCATCGAGAACCTCCACGGACTGTC 19022
Qy 14241 CATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCA 14300
Db 19023 GATCGCCGCCGTCAACGGCCCCACCGCCACCGTGGTTTCGGGCGACCCCACCCAGATCCA 19082
Qy 14301 ACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCA 14360
Db 19083 AGAGCTCGCTCAGGCGTGTGAGGCCGACGGGGTCCGCGCACGGATCATCCCCGTCGACTA 19142

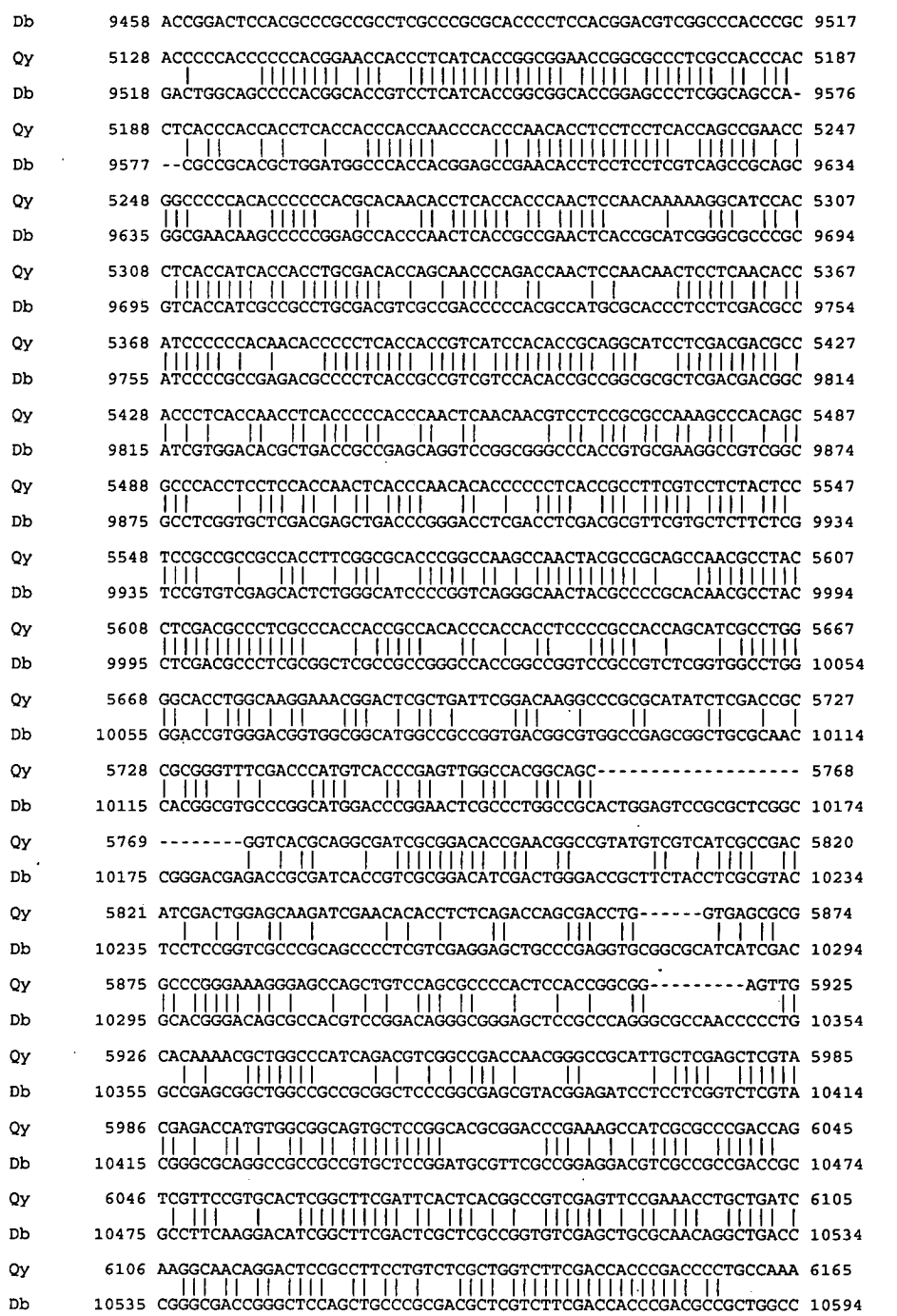

Qy 10285 GAGCATCCGGACCGGTTTGTCCTCCTCGACCTCGACACCGACACCGGCACCGACCTCGAC 10344
Db 14882 CACCTCGCGGGCACCCGCCTGCTCCTGGTCACCCGCGGTGCGGTGCGGGACCCCGAGGGG 14941
Qy 10345 ACCGGTGCTGGTGCTGGTTGGGGC-----GTGGATGGTGGGCGTGTGGCGGCGGTGGTGG 10399
Db 14942 TCCGGCGCCGACGATGGCGGCGAGGACCTGTCGCACGCGGCCGCCTGGGGTCTCGTACGG 15001
Qy 10400 CGTGTGGTGAGCCGCAGTTGGCGGTGCGTGGGGAGCGGTTGCTGGCCGCACGCCTGAAAC 10459
Db 15002 ACCGCGCAGACCGAGAACCCCGGCCGCTTCGGCCTTCTCGACCTGGCCGACGACGCCTCG 15061
Qy 10460 GACTTGAGTCATCCGGTGATGTTCCAGCCCAGCGGTCCGGTGACACACGAGCCCGGCGGT 10519
Db 15062 TCGTACCGGACCCTGCCGTCGGTGCTCTCCGACGCGGGCCTGCGCGACGAACCGCAGCTC 15121
Qy 10520 CCGACGTGCCTGCCCAGCGCTCCGGTGGCGTGCCTGCTCGGCGGTCGGTTGATGTATCGG 10579
Db 15122 GCCCTGCACGACGGCACCATCAGGCTGGCCCGCCTGGCCTCCGTCCGG--CCCGAGACCG 15179
Qy 10580 GTCGGGAGGTGTTGCCGTGGTTGTCGGGTGGGTCGGTGTTGGTGACGGGTGGGACGGGTG 10639
Db 15180 GCACCGCCGCACCGGCGCTCGCCCCGGAGGGCACGGTCCTGCTGACCGGCGGCACCGGCG 15239
Qy 10640 TGCTGGGTGCGGCGGTGGCGCGGCATCTGGCTGGTGTGTGTGGGGTGCGGGATCTGCTGT 10699
Db 15240 GCCTGGGCGGACTGGTCGCCCGGCACGTGGTGGGCGAGTGGGGCGTACGACGCCTGCTGC 15299
Qy 10700 TGGTGAGCCGGCGTGGTCCGGATGCTCCGGGTGCGGAGGGTCTGCGGGCGGAGCTGGCCG 10759
Db 15300 TGGTGAGCCGGCGGGGCACGGACGCCCCGGGCGCCGACGAGCTCGTGCACGAGCTGGAGG 15359
Qy 10760 CGTTGGGGGCGGAGGTGCGGATTGTTGCGTGTGATGTGGGGGAGCGGCGGGAGGTGGTCC 10819
Db 15360 CCCTGGGAGCCGACGTCTCGGTGGCCGCGTGCGACGTCGCCGACCGCGAAGCCCTCACCG 15419
Qy 10820 GGCTGCTGGAGGGTGTTCCTGCCGGGTGTCCGCTGACGGGTGTCGTGCATGCGGCTGGTG 10879
Db 15420 CCGTACTCGACGCCATCCCCGCCGAACACCCGCTCACCGCGGTCGTCCACACGGCAGGCG 15479
Qy 10880 TGCTGGACGATGCGACGATCGCCTCTCTCACGCCCGAGCGGCTGGGCACGGTGTTCGCGG 10939
Db 15480 TCCTCTCCGACGGCACCCTCCCGTCCATGACGACGGAGGACGTGGAACACGTACTGCGGC 15539
Qy 10940 CCAAGGTGGATGCCGCTCTTTTGCTGGATGAGCT------GACGCGGGGTATGGAGCTGT 10993
Db 15540 CCAAGGTCGACGCCGCGTTCCTCCTCGACGAACTCACCTCGACGCCCGCATACGACCTGG 15599
Qy 10994 CGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATCCTGGGGTCGGCCGGGCAGGGCAACT 11053
Db 15600 CAGCGTTCGTCATGTTCTCCTCCGCCGCCGCCGTCTTCGGTGGCGCGGGGCAGGGCGCCT 15659
Qy 11054 ACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCGTACCGGCGGCGGGCGGCGGGTCTGC 11113
Db 15660 ACGCCGCCGCCAACGCCACCCTCGACGCCCTCGCCTGGCGCCGCCGGGCAGCCGGACTCC 15719
Qy 11114 CGGGGGTGTCGCTGGCGTGGGGGCTGTGGGAAGAGGCCAGCGGGATGACCGGGCACCTGG 11173
Db 15720 CCGCCCTCTCCCTCGGCTGGGGCCTCTGGGCCGAGACCAGCGGCATGACCGGCGAGCTCG 15779
Qy 11174 CCGGCACCGACCACCGGCGCATCATCCGTTCCGGTCTGCATCCCATGTCGACCCCGGACG 11233
Db 15780 GCCAGGCGGACCTGCGCCGGATGAGCCGCGCGGGCATCGGCGGGATCAGCGACGCCGAGG 15839
Qy 11234 CACTGGCCCTCTTCGATGCGGCCCTGGCTCTGGACCGGCCGGTCCTGCTGCCCGCCGACC 11293
Db 15840 GCATCGCGCTCCTCGACGCCGCCCTCCGCGACGACCGCCACCCGGTCCTGCTGCCCCTGC 15899
Qy 11294 TGCGTCCCGCCCCGCCCCTGCCG---------- 11316
Db 15900 GGCTCGACGCCGCCGGGCTGCGGGACGCGGCCGGGAACGACCCGGCCGGAATCCCGGCGC 15959
Qy 11317 ----------CCCCTGCTGCAGGACCTCCTGCCCGCCACCCGCCGCCGCACCACCCG 11363
Db 15960 TCTTCCGGGACGTCGTCGGCGCCAGGACCGTCCGGGCCCGGCCGTCCGCGGCCTCCGCCT 16019
Qy 11364 CACCACCACTACCGGTGGTGCGGACAACGGC---------- 11394
Db 16020 CGACGACAGCCGGGACGGCCGGCACGCCGGGGACGGCGGACGGCGCGGCGGAAACGGCGG 16079
Qy 11395 --GCCCAGCTGCACGCCCGGCTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCC 11452
Db 16080 CGGTCACGCTCGCCGACCGGGCCGCCACCGTGGACGGGCCCGCACGGCAGCGCCTGCTGC 16139
Qy 11453 TCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCC 11512
Db 16140 TCGAGTTCGTCGTCGGCGAGGTCGCCGAAGTACTCGGCCACGCCCGCGGTCACCGGATCG 16199
Qy 11513 CCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCA 11572
Db 16200 ACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGACCGCCGTCGAACTCCGCA 16259
Qy 11573 ACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCA 11632
Db 16260 ACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGCGACCCTGGTCTTCGACCACCCAA 16319
Qy 11633 ACCCCACCACCCTCACCCACCACCTCCACACACAACTCC---------- 11671
Db 16320 GCCCGGCGGCACTCGCCTCCCACCTGGACGCCGAGCTGCCGCGCGGCGCCTCGGACCAGG 16379
Qy 11672 ----------AGCCACAACCGGACAACGCTGTCGC---------- 11696
Db 16380 ACGGAGCCGGGAACCGGAACGGGAACGAGAACGGGACGACGGCGTCCCGGAGCACCGCCG 16439
Qy 11697 -------CCCCGTGTTGGCGGAGCTCGACAAACTCGAATCCGCCCTCTCCGCCCTCGACA 11749
Db 16440 AGACGGACGCGCTGCTGGCACAACTGACCCGCCTGGAAGGCGCCTTGGTGCTGACGGGCC 16499
Qy 11750 AAACCGACAGCGCC---AGCGAAAGAGTCACCCTGCGGCTGAAGTCACTCATGTTGAGG- 11805
Db 16500 TCTCGGACGCCCCCGGGAGCGAAGAAGTCCTGGAGCACCTGCGGTCCCTGCGCTCGATGG 16559
Qy 11806 ----------TGGAACGCACCCCAGCATCCGACAGCCGAAAGCGCTGATG 11845
Db 16560 TCACGGGCGAGACCGGGACCGGGACCGCGTCCGGAGCCCCGGACGGCGCCGGGTCCGGCG 16619
Qy 11846 ACGACGA---------- 11852
Db 16620 CCGAGGACCGGCCCTGGGCGGCCGGGGACGGAGCCGGGGGCGGGAGTGAGGACGGCGCGG 16679
Qy 11853 -------GAAGTTCACATCGGCAACAGAGGCTGAGATTTTCAAATTCATTGACAACGACC 11905
Db 16680 GAGTGCCGGACTTCATGAACGCCTCGGCCGAGGAACTCTTCGGCCTCCTCGACCAGGACC 16739
Qy 11906 TCGGCCTGTCCTGAACCGGACGCCTGCCACTCCGCCCGTATCCG---------- 11949
Db 16740 CCAGCACGGACTGATCCCTGCCGCACGGTCGCCTCCCGCCCCGGACCCCGTCCCGGGCAC 16799
Qy 11950 ----------CTGGGCCCTGCTAGGACGTGAAT 11972
Db 16800 CTCGACTCGAATCACTTCATGCGCGCCTCGGGCGCCTCCAGGAACTCAAGGGGACAGCGT 16859
Qy 11973 GCAATTGGCGAATGAAGCGAAGCTCCTGGAATACCTCAAGCGCGTCACTGCGGACCTGGA 12032
Db 16860 GTCCACGGTGAACGAAGAGAAGTACCTCGACTACCTGCGTCGTGCCACGGCGGACCTCCA 16919
Qy 12033 CCGCACTCGCCGTCGCCTGTACGAGGTGGTCGAGCGTGAGCAGGAGCCGATCGCGATTGT 12092
Db 16920 CGAGGCCCGTGGCCGCCTCCGCGAGCTGGAGGCGAAGGCGGGCGAGCCGGTGGCGATCGT 16979
Qy 12093 GGGGATGGCGTGTCGTTACCCAGGCGGGGCGACGTCACCCACGCGACTGTGGCATCTCGT 12152
Db 16980 CGGCATGGCCTGCCGCCTGCCCGGCGGCGTCGCCTCGCCCGAGGACCTGTGGCGGCTGGT 17039
Qy 12153 CAAGTCCCAGACGGACGCTATCGGGGAGTTCCCGACCGACCGTGGATGGAACCTGGAGCA 12212

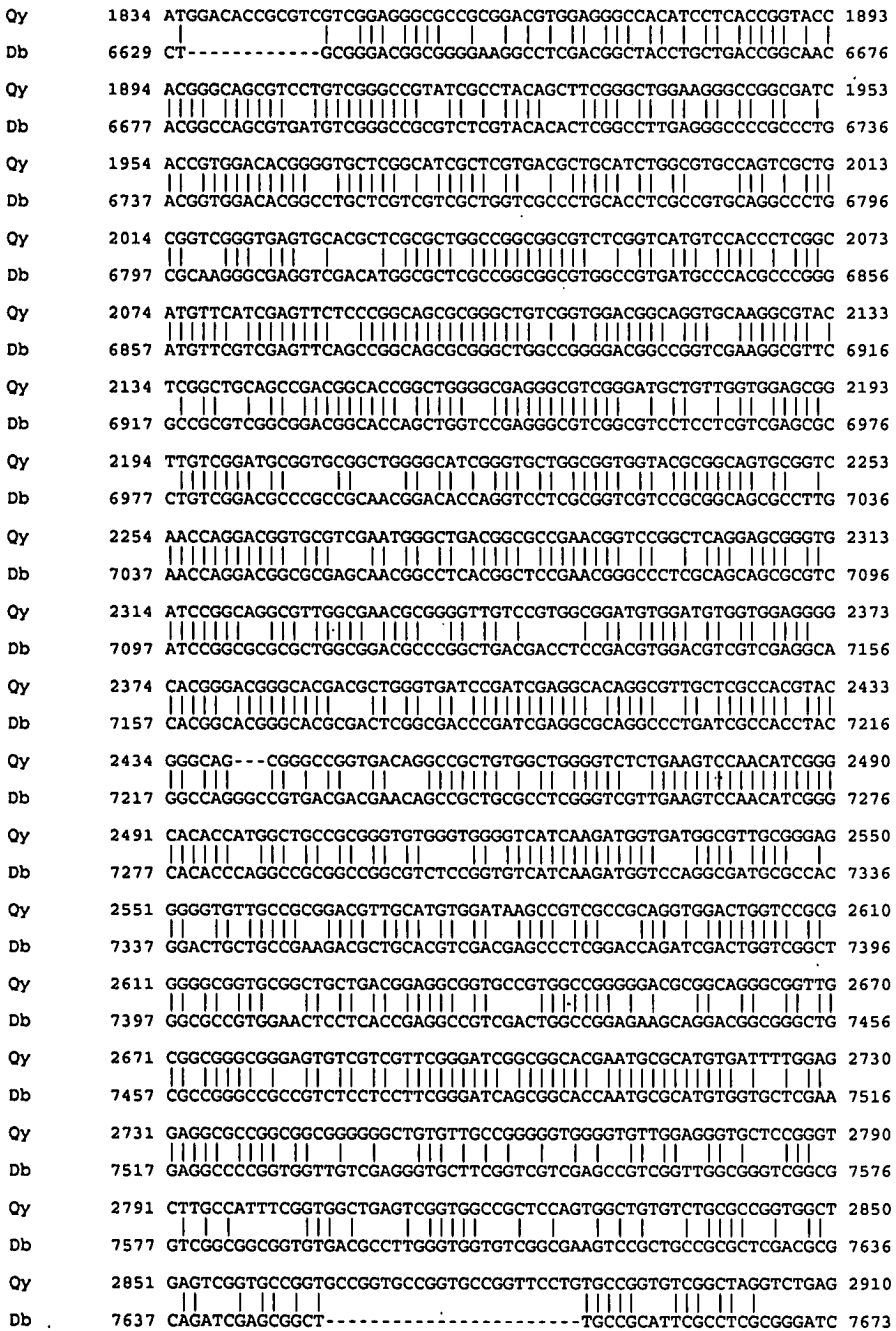

Db 12719 CTCGGTCACTCGGTCGGCGAGATCGCCGCCGCGCACGTCGCCGGTGTGTTCTCGCTCGCC 12778
Qy 8251 GACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCC---GGC 8307
Db 12779 GACGCCGCCCGCCTGGTCGCCGCGCGCGGCCGGCTCATGCAGGAGCTGCCCGCCGGTGGC 12838
Qy 8308 ACCATGACCACCCTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAA 8367
Db 12839 GCGATGCTCGCCGTCCAGGCCGCGGAGGACGAGATCCGCGTGTGGCTGGAGACGGAGGAG 12898
Qy 8368 AACGAC---------CTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGC 8418
Db 12899 CGGTACGCGGGACGTCTGGACGTCGCCGCCGTCAACGGCCCCGAGGCCGCCGTCCTGTCC 12958
Qy 8419 GGCACCCCCCACACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACC 8478
Db 12959 GGCGACGCGGACGCGGCGCGGGAGGCGGAGGCGTACTGGTCCGGGCTCGGCCGCAGGACC 13018
Qy 8479 AAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAA 8538
Db 13019 CGCGCGCTGCGGGTCAGCCACGCCTTCCACTCCGCGCACATGGACGGCATGCTCGACGGG 13078
Qy 8539 CTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCC 8598
Db 13079 TTCCGCGCCGTCCTGGAGACGGTGGAGTTCCGGCGCCCCTCCCTGACCGTGGTCTCGAAC 13138
Qy 8599 AACA----------CCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAA 8646
Db 13139 GTCACCGGCCTGGCCGCCGGCCCGGACGACCTGTGCGACCCCGAGTACTGGGTCCGGCAC 13198
Qy 8647 GCCCGCAACACCGTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACC 8706
Db 13199 GTCCGCGGCACCGTCCGCTTCCTCGACGGCGTCCGTGTCCTGCGCGACCTCGGCGTGCGG 13258
Qy 8707 ACCTACATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCC 8766
Db 13259 ACCTGCCTGGAGCTGGGCCCCGACGGGGTCCTCACCGCCATGGCGGCCGACGGCCTCGCG 13318
Qy 8767 AACCCCCCCACCACCACCCTCACCCTCACCCAC---------- 8799
Db 13319 GACACCCCCGCGGATTCCGCTGCCGGCTCCCCCGTCGGCTCTCCCGCCGGCTCTCCCGCC 13378
Qy 8800 ----------CCCCACCACCACCCCCAAACCCACCTCCTCACCAAC 8835
Db 13379 GACTCCGCCGCCGGCGCGCTCCGGCCCCGGCCGCTGCTCGTGGCGCTGCTGCGCCGCAAG 13438
Qy 8836 CTCGCCAAAACCACCACCACCTGGCACCCCCACCACTACACCCA---CCACGACAACCAA 8892
Db 13439 CGGTCGGAGACCGAGACCGTCGCGGACGCCCTCGGCAGGGCGCACGCCCACGGCACCGGA 13498
Qy 8893 CCCCACACCCACAC----------CCACCTCGACCTCCCCACCTAC 8928
Db 13499 CCCGACTGGCACGCCTGGTTCGCCGGCTCCGGGGCGCACCGCGTGGACCTGCCCACGTAC 13558
Qy 8929 CCCTTCCAACACCACCACTACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCA 8988
Db 13559 TCCTTCCGGCGCGACCGCTACTGGCTGGACGCCCCGGCGGCCGACACCGCGGTGGACACC 13618
Qy 8989 GCAG---CCGGACTCGACCCCACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCG 9045
Db 13619 GCCGGCCTCGGTCTCGGCACCGCCGACCACCCGCTGCTCGGCGCCGTGGTCAGCCTTCCG 13678
Qy 9046 ACTGACGGTGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCT 9105
Db 13679 GACCGGGACGGCCTGCTGCTCACCGGCCGCCTCTCCCTGCGCACCCACCCGTGGCTCGCG 13738
Qy 9106 GACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTT 9165
Db 13739 GACCACGCCGTCCTGGGGAGCGTCCTGCTCCCCGGCGCCGCGATGGTCGAACTCGCCGCG 13798
Qy 9166 CATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTG 9225

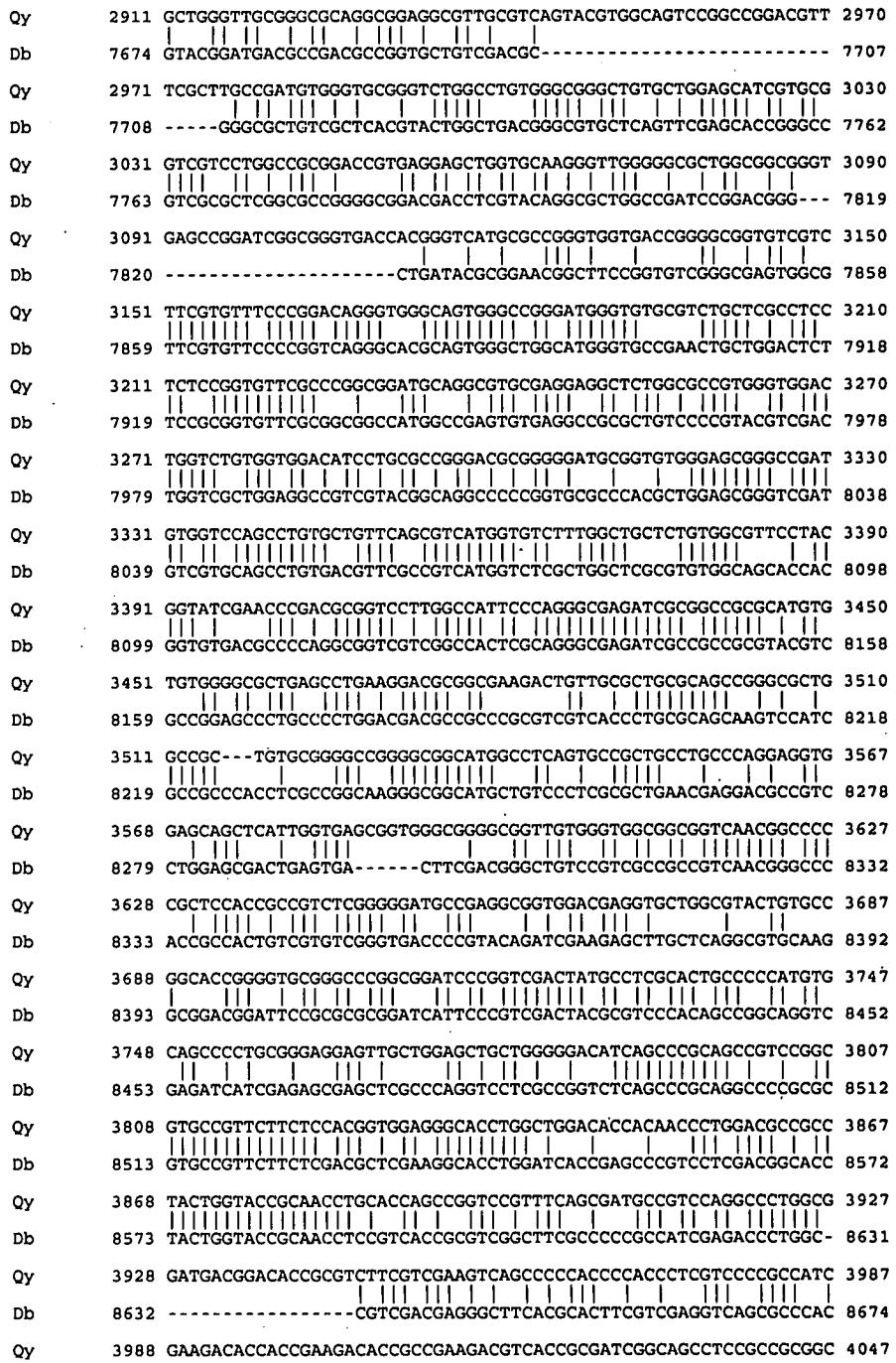

Db 13799 CACGCTGCGGAGTCCGCCGGTCTGCGTGACGTGCGGGAGCTGACCCTCCTTGAACCGCTG 13858
Qy 9226 GTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGG---- 9281
Db 13859 GTACTGCCCGAGCACGGTGGCGTCGAGCTGCGCGTGACGGTCGGGGCGCCGGCCGGAGAG 13918
Qy 9282 ---GGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTTGTGG--TGGG 9336
Db 13919 CCCGGTGGCGAGTCGGCCGGGGACGGCGCACGGCCCGTCTCCCTCCACTCGCGGCTCGCC 13978
Qy 9337 GGTGGTGCGTCGGGTGGGGTGTGGACGTGTCATGCCTCGGGGGTGCTGGTTGAGGCTGCT 9396
Db 13979 GACGCGCCCGCCGGTACCGCCTGGTCCTGCCACGCGACCGGTCTGCTGGCCACCGACCGG 14038
Qy 9397 GCTGGTGGTGTGGTGGTGGATGGTCTGGCGGGGGTGTGGCCGCCGCGGGGTGCGGTGGCG 9456
Db 14039 CCCGAGCTTCCCGTCGCGCCCGACCGTGCGGCCATGTGGCCGCCGCAGGGCGCCGAGGAG 14098
Qy 9457 GTGGATGTCGATGGTGTCCGTGACCGTTTGGCTGGGGCTGGTTGTGTTTTGGGGCCGGTG 9516
Db 14099 GTGCCGCTCGACGGTCTCTACGAGCGGCTCGACGGGAACGGCCTCGCCTTCGGTCCGCTG 14158
Qy 9517 TTTTCGGGGCTGCGTGCGGTGTGGCGTGATGGGGGGGATTTGCTGGCTGAGGTGTGTCTG 9576
Db 14159 TTCCAGGGGCTGAACGCGGTGTGGCGGTACGAGGGTGAGGTCTTCGCCGACATCGCGCTC 14218
Qy 9577 CC----------GGAGGAGGCGTGGGGTGATGCGGCTGGTTTTGGGCTGCATCCGGCG 9624
Db 14219 CCCGCCACCACGAATGCGACCGCGCCCGCGACCGCGAACGGCGGCGGGAGTGCGGCGGCG 14278
Qy 9625 TTGCTGGATGGTGTGGTCCAGCCGTTGTCGGTGTTGCTTCCGGGTGGGACGGGGTTTGGG 9684
Db 14279 GCCCCCTACGGCATCCACCCCGCCCTGCTCGACGCTTCGCTGCACGCCATCGCGGTCGGC 14338
Qy 9685 GAGGGGGCGGGGTTCGGGGAGGGTGTTCGGGTGCCGGCTGTGTGGGGTGGTGTGTCGCTT 9744
Db 14339 GGTCTCGTCGACGAGCCCGAGCTCGTCCGCGTCCCCTTCCACTGGAGCGGTGTCACCGTG 14398
Qy 9745 CACCGGGCGGGTGTGACCGGTGTGCGGGTGCGTGTGTCGGCTGTCGGGCGGGGCGGCGGG 9804
Db 14399 CACGCGGCCGGTGCCGCGGCGGCCCGGGTCCGTCTCGCCTC----------CGCGGGG 14446
Qy 9805 CGTGAGGCGGTGTCGGTCGTGGTCGGGGATGAGGCGGGTGTGCCGGTGGCGTCGGTCGAT 9864
Db 14447 ACGGACGCCGTCTCGCTGTCCCTGACGGACGGCGAGGGACGCCCGCTGGTCTCCGTGGAA 14506
Qy 9865 CGTCTTGAGTTGCGGCCTGTGGATATGGGTCAGTTGCGTGCTGTCTCGGTTTCGGCGGGG 9924
Db 14507 CGGCTCACGCTGCGCCCGGTCACCGCCGATCAGGCGGCGGC----------GAGCCG 14553
Qy 9925 CGGCGGGGTTCGCTGTATGCGGTGCAGTGGGCTGAGGTGGGTCCTGTGCCGGTGTGTGGG 9984
Db 14554 CGTCGGCGGGCTGATGCACCGGGTGGCCTGGCGTCCGTACGCCCTCGCCTCGTCCGGCGA 14613
Qy 9985 CAGGCGTGGGCGTGGCACGAGGACGTGGGTGAGAGCGGTGGTGGGCCTGTGCCGGGGGTG 10044
Db 14614 ACAGGACCCGCACGCCACTTCGTACGGGCCGACCGCCGTCCTCGGCAAG---------- 14662
Qy 10045 GTGGTGTTGCGGTGCCCGGATGCCGGTGCCGGTGGCGGTGGCGGTGGCGGTGGTGGCGGT 10104
Db 14663 ----------GACGAGCTGAAGGTCGCCGCCGCCCTGGAGTCCGCGGGCGTCGAAGT 14709
Qy 10105 GGTGTGGGTGAGGTTGTTGGTGGGGTGTTGGGTGTGGTGCAGGGGTGGCTGGGGCTGGAG 10164
Db 14710 --------CGGGCTCTACCCCGACCTGGCCGCGCTGTCCCAGGACGTGGCGGCCGGCGCC 14761
Qy 10165 CGGTTTGCGGGTTCGCGGCTGGTGGTGGTGACCCGGGGTGCGGTGGTGGCCGGCCCCGGAG 10224
Db 14762 CCGGCGCCCCGTACCGTCCTTGCGCCGCTGCCCGCGGGTCCCGCCGACGGCGGCGCGGAG 14821
Qy 10225 GACGGCCCGGTGGATGTGGTGGGTGCGTCGGTGTGGGGGCTGGTGCGTTCGGCGCAGGCT 10284
Db 14822 GGTGTACGGGGCACGGTGGCCCGGACGCTGGAGCTGCTCCAGGCCTGGCTGGCCGACGAG 14881


Qy 6166 CTCGCCGTACACCTGCAGAACCAACTGCGGGGCACAGCAGCGGAGTCGGC---------- 6215
Db 10595 CTCGTGTCGCTGCTCCGCAGCGAGTTCCTCGGTGACGAGGAGACGGCGGACGCCCGGCGG 10654
Qy 6216 ----------TCCTTCAGCGGCAGCCGTTACCGCCGAGGCTTCT 6249
Db 10655 TCCGCGGCGCTGCCCGCGACTGTCGGTGCCGGTGCCGGCGCCGGCGCCGGCACCGATGCC 10714
Qy 6250 GTCACCGAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCG 6309
Db 10715 GACGACGATCCGATCGCGATCGTCGCGATGAGCTGCCGCTACCCCGGTGACATCCGCAGC 10774
Qy 6310 GCGGACGACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACC 6369
Db 10775 CCGGAGGACCTGTGGCGGATGCTGTCCGAGGGCGGCGAGGGCATCACGCCGTTCCCCACC 10834
Qy 6370 GACCGCGGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGC 6429
Db 10835 GACCGCGGCTGGGACCTCGACGGCCTGTACGACGCCGACCCGGACGCGCTCGGCAGGGCG 10894
Qy 6430 TACACCCGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGC 6489
Db 10895 TACGTCCGCGAGGGCGGGTTCCTGCACGACGCGGCCGAGTTCGACGCGGAGTTCTTCGGC 10954
Qy 6490 ATCAGCCCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCC 6549
Db 10955 GTCTCGCCGCGCGAGGCGCTGGCCATGGACCCGCAGCAGCGGATGCTCCTGACGACGTCC 11014
Qy 6550 TGGGAAACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGA 6609
Db 11015 TGGGAGGCCTTCGAGCGGGCCGGCATCGAGCCGGCATCGCTGCGCGGCAGCAGCACCGGT 11074
Qy 6610 GTCTTCACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCA 6669
Db 11075 GTCTTCATCGGCCTCTCCTACCAGGACTACGCGGCCCGCGTCCCGAACGCCCCGCGTGGC 11134
Qy 6670 ACCGATGGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTAC 6729
Db 11135 GTGGAGGGTTACCTGCTGACCGGCAGCACGCCGAGCGTCGCGTCGGGCCGTATCGCGTAC 11194
Qy 6730 ACGTTTGGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTG 6789
Db 11195 ACCTTCGGTCTCGAAGGGCCCGCGACGACCGTCGACACCGCCTGCTCGTCGTCGCTGACC 11254
Qy 6790 GCTTTGCATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGG 6849
Db 11255 GCCCTGCACCTGGCGGTGCGGGCGCTGCGCAGCGGCGAGTGCACGATGGCGCTCGCCGGT 11314
Qy 6850 GGTGTGACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTG 6909
Db 11315 GGCGTGGCGATGATGGCGACCCCGCACATGTTCGTGGAGTTCAGCCGTCAGCGGGCGCTC 11374
Qy 6910 GCCGCGGACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAG 6969
Db 11375 GCCCCGGACGGCCGCAGCAAGGCCTTCTCGGCGGACGCCGACGGGTTCGGCGCCGCGGAG 11434
Qy 6970 GGTGTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTC 7029
Db 11435 GGCGTCGGCCTGCTGCTCGTGGAGCGGCTCTCGGACGCGCGGCGCAACGGTCACCCGGTG 11494
Qy 7030 CTGGCCGTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCG 7089
Db 11495 CTCGCCGTGGTCCGCGGTACCGCCGTCAACCAGGACGGCGCCAGCAACGGGCTGACCGCG 11554
Qy 7090 CCCAACGGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCG 7149
Db 11555 CCCAACGGACCCTCGCAGCAGCGGGTGATCCGGCAGGCGCTCGCCGACGCCCGGCTGGCA 11614
Qy 7150 GCCGGTGATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATC 7209
Db 11615 CCCGGCGACATCGACGCCGTCGAGACGCACGGCACGGGAACCTCGCTGGGCGACCCCATC 11674
Qy 7210 GAGGCCCAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGG 7269
Db 11675 GAGGCCCAGGGCCTCCAGGCCACGTACGGCAAGGAGCGGCCCGCGGAACGGCCGCTCGCC 11734
Qy 7270 CTGGGCTCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTG 7329
Db 11735 ATCGGCTCCGTGAAGTCCAACATCGGACACACCCAGGCCGCGGCCGGTGCGGCGGGCATC 11794
Qy 7330 ATCAAGATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAG 7389
Db 11795 ATCAAGATGGTCCTCGCGATGCGCCACGGCACCCTGCCGAAGACCCTCCACGCCGACGAG 11854
Qy 7390 CCGTCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCC 7449
Db 11855 CCGAGCCCGCACGTCGACTGGGCGAACAGCGGCCTGGCCCTCGTCACCGAGCCGATCGAC 11914
Qy 7450 TGGCCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGC 7509
Db 11915 TGGCCGGCCGGCACCGGTCCG---CGCCGCGCCGCCGTCTCCTCCTTCGGCATCAGCGGG 11971
Qy 7510 ACCAACGCCCACGTCATCCTCGAAGAAGCACCCGCCGACGACGTT---------- 7554
Db 11972 ACGAACGCGCACGTCGTGCTGGAGCAGGCGCCGGATGCTGCTGGTGAGGTGCTTGGGGCC 12031
Qy 7555 ----------CCGGGGGGACCACCC 7569
Db 12032 GATGAGGTGCCTGAGGTGTCTGAGACGGTAGCGATGGCTGGGACGGCTGGGACCTCCGAG 12091
Qy 7570 GCCGGCGAGGGTGACGCGGGCAGCGACGATGAGGCTGCTGCCGGCAGTCCTGGGGTGT-- 7627
Db 12092 GTCGCTGAGGGCTCTGAGGCCTCCGAGGCCCCCGCGGCCCCCGGCAGCCGTGAGGCGTCC 12151
Qy 7628 ----------GGCCGTGGCTGGTGTCGGCCAAGTCGCAGCCGGCCCTGCGCGCCCAG 7674
Db 12152 CTCCCCGGGCACCTGCCCTGGGTGCTGTCCGCCAAGGACGAGCAGTCGCTGCGCGGCCAG 12211
Qy 7675 GCCCAGGCCCTGCACGCCCACCTCACCGACCACCCCGGCCTCGA---------- 7718
Db 12212 GCCGCCGCCCTGCACGCGTGGCTGTCCGAGCCCGCCGCCGACCTGTCGGACGCGGACGGA 12271
Qy 7719 --------CCTCGCGGATGTCGGATACACCCTCGCCCACGCCCGCGCCGTGTTCGACCAC 7770
Db 12272 CCGGCCCGCCTGCGGGACGTCGGGTACACGCTCGCCACGAGCCGTACCGCCTTCGCGCAC 12331
Qy 7771 CGCGCCACCCTCATCGCCGCGGACCGCGACACGTTCCTGCAAGCACTCCAGGCACTCGCC 7830
Db 12332 CGCGCCGCCGTGACCGCCGCCGACCGGGACGGGTTCCTGGACGGGCTGGCCACGCTGGCC 12391
Qy 7831 GCAGGCGAGCCCCACCCCGCCGTCATCCACAGCAGCGCCCCGGGCGGGACCGGGACCGGG 7890
Db 12392 CAGGGCGGCACCTCGGCCCACGTCCACCTGGACACCGCCCGGGACGGCACC---------- 12442
Qy 7891 GAGGCCGCAGGAAAGACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATG 7950
Db 12443 ----------ACCGCGTTCCTCTTCACCGGCCAGGGCAGTCAGCGCCCCGGCGCC 12487
Qy 7951 GCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACC 8010
Db 12488 GGCCGTGAGCTGTACGACCGGCACCCCGTCTTCGCCCGGGCGCTCGACGAGATCTGCGCC 12547
Qy 8011 CACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGAC 8070
Db 12548 CACCTCGACGGTCACCTCGAACTGCCCCTGCTCGACGTGAT------GTTCGCGGCCGAG 12601
Qy 8071 AACGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCC 8130
Db 12602 GGCAGCGCGGAGGCCGCGCTGCTCGACGAGACGCGGTACACGCAGTGCGCGCTGTTCGCC 12661
Qy 8131 TTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTAC 8190
Db 12662 CTGGAGGTCGCGCTCTTCCGGCTCGTC---GAGAGCTGGGGCATGCGGCCGGCCGCACTG 12718
Qy 8191 GCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACC 8250


Db 8675 CCCGTCCTCACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGCACCCTCCGTCGCGAA 8734
Qy 4048 GACAACGACACCCGCCGCTTCCTCACCGCCCTCGCCCACACCCATACCACCGGCATCGGC 4107
Db 8735 CAGGGAGGCCAAGAGCGTCTGGTCACCTCGCTCGCCGAGGCGTGGGTCAACGGGCTTCCC 8794
Qy 4108 ACACCCACCACCTGGCACCACCACTACACCCACCACCACACCCACCCCCACCCCCACACG 4167
Db 8795 GTGGCATGGACTTCGCTCCTGCCCGCCACGGCCTCCCGCCC---------- 8835
Qy 4168 CACCTCGACCTGCCCACCTACCCCTTCCAACACCAGCACTACTGGCTCGAGAGCTCACAG 4227
Db 8836 -----CGGTCTGCCCACCTACGCCTTCCAGGCCGAGCGCTACTGGCTCGAGAACACTCCC 8890
Qy 4228 CCGGGTGCCGGATCCGGTTCGGGTGCCGGTGCCGGTTCGGGTGCCGGTTCCGGGCGGGCA 4287
Db 8891 GCCGC----------CCTGGCCACCG 8906
Qy 4288 GGGACTGCGGGCGGGACGGCAGAGGTGGAGTCGCGGTTCTGGGACGCGGTGGCCCGCCAG 4347
Db 8907 GCGACGACTGGCGCTACCGCATCGACTG-GAAGCGCCTCCCGGCCGCCGAGGGGTCCGAG 8965
Qy 4348 GACCTGGAAACGGTCGCGACCACACTCGCCGTGCCCCCCTCCGCCGGCCTGGACACGGTG 4407
Db 8966 CGCACCGGCCTGTCCGGCCGCTGGCTCGCCGTCACGCCGGAGGACCACTCCGCGCAGGCC 9025
Qy 4408 GTGCCCGCACTCTCCGCCTGGCACCGCCACCAACACGACCAAGCCCGCATCAACACCTGG 4467
Db 9026 GCCGCCGTGCTCACCGCGCTGGTCGACGCCGGGGCGAAGGTCGAGGTGCTGACGGCCGGG 9085
Qy 4468 ACCTACCAGGAAACCTGGAAACCCCTCACCCTCCCCACCACCCACCAACCCCACCAAACC 4527
Db 9086 GCGGACGACGACCGTGAGGCCCTCGCCGCCCGGCTCACCGCACTGACG---------- 9133
Qy 4528 TGGCTCATCGCCATCCCCGAAACCCAGACCCACCACCCCCACATCACCAACATCCTCACC 4587
Db 9134 ---------- 9133
Qy 4588 AACCTCCACCACCACGGCATCACCCCCATCCCCCTCACCCTCAACCACACCCACACCAAC 4647
Db 9134 ---------- 9133
Qy 4648 CCCCAACACCTCCACCACACCCTCCACCACACCCGACAACAAGCCCAAAACCACACCACC 4707
Db 9134 ----------ACCGGT 9139
Qy 4708 GGAGCCATCACCGGCCTGCTCTCCCTCCTCGCCCTCGACGAAACACCCCACCCCCACCAC 4767
Db 9140 GACGGCTTCACCGGCGTGGTCTCGCTCCTCG---------- 9170
Qy 4768 CCCCACACACCCACCGGCACCCTCCTCAACCTCACCCTCACCCAAACCCACACCCAAACC 4827
Db 9171 ----------ACGGACTCGTACCGCAGGTCGCCTGGGTCCAGGCGCTCGGCGACGCC 9217
Qy 4828 CACCCACCAACCCCCCTCTGGTACGCCACCACCAACGCCACCACCACCCACCCCAACGAC 4887
Db 9218 GGAATCAAGGCGCCCCTGTGGTCCGTCACCCAGGGCGCGGTCTCCGTCGGACGTCTCGAC 9277
Qy 4888 CCCCTCACACACCCCACCCAAGCCCAAACCTGGGGACTCGCCCGCACCACCCTCCTCGAA 4947
Db 9278 ACCCCCGCCGACCCCGACCGGGCCATGCTCTGGGGCCTCGGCCGCGTCGTCGCCCTTGAG 9337
Qy 4948 CACCCCACCCACACCGCCGGAATCATCGACCTCCCCACCACCCCCACCCCCCACACCCTC 5007
Db 9338 CACCCCGAACGCTGGGCCGGCCTCGTCGACCTCCCCGCCCAGCCCGATGCCGCCGCCCTC 9397
Qy 5008 CAGCACCTCACCCAAACCCTCACCCAACCCCACCACCAAACCCAACTCGCCATCCGCACC 5067
Db 9398 GCCCACCTCGTCACCGCACTCTCCGGCGCCACCGGCGAGGACCAGATCGCCATCCGCACC 9457
Qy 5068 ACCGGCACCCACACCCGCCGCCTCACCCCCACCACCCTCACCCCCACACACCAACCACCC 5127


Db 9458 ACCGGACTCCACGCCCGCCGCCTCGCCCGCGCACCCCTCCACGGACGTCGGCCCACCCGC 9517
Qy 5128 ACCCCCACCCCCCCACGGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCAC 5187
Db 9518 GACTGGCAGCCCCACGGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCA- 9576
Qy 5188 CTCACCCACCACCTCACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACC 5247
Db 9577 --CGCCGCACGCTGGATGGCCCACCACGGAGCCGAACACCTCCTCCTCGTCAGCCGCAGC 9634
Qy 5248 GGCCCCCACACCCCCCACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCAC 5307
Db 9635 GGCGAACAAGCCCCCGGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGC 9694
Qy 5308 CTCACCATCACCACCTGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACC 5367
Db 9695 GTCACCATCGCCGCCTGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCC 9754
Qy 5368 ATCCCCCCACAACACCCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCC 5427
Db 9755 ATCCCCGCCGAGACGCCCCTCACCGCCGTCGTCCACACCGCCGGCGCGCTCGACGACGGC 9814
Qy 5428 ACCCTCACCAACCTCACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGC 5487
Db 9815 ATCGTGGACACGCTGACCGCCGAGCAGGTCCGGCGGGCCCACCGTGCGAAGGCCGTCGGC 9874
Qy 5488 GCCCACCTCCTCCACCAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCC 5547
Db 9875 GCCTCGGTGCTCGACGAGCTGACCCGGGACCTCGACCTCGACGCGTTCGTGCTCTTCTCG 9934
Qy 5548 TCCGCCGCCGCCACCTTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTAC 5607
Db 9935 TCCGTGTCGAGCACTCTGGGCATCCCCGGTCAGGGCAACTACGCCCCGCACAACGCCTAC 9994
Qy 5608 CTCGACGCCCTCGCCCACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGG 5667
Db 9995 CTCGACGCCCTCGCGGCTCGCCGCCGGGCCACCGGCCGGTCCGCCGTCTCGGTGGCCTGG 10054
Qy 5668 GGCACCTGGCAAGGAAACGGACTCGCTGATTCGGACAAGGCCCGCGCATATCTCGACCGC 5727
Db 10055 GGACCGTGGGACGGTGGCGGCATGGCCGCCGGTGACGGCGTGGCCGAGCGGCTGCGCAAC 10114
Qy 5728 CGCGGGTTTCGACCCATGTCACCCGAGTTGGCCACGGCAGC---------- 5768
Db 10115 CACGGCGTGCCCGGCATGGACCCGGAACTCGCCCTGGCCGCACTGGAGTCCGCGCTCGGC 10174
Qy 5769 --------GGTCACGCAGGCGATCGCGGACACCGAACGGCCGTATGTCGTCATCGCCGAC 5820
Db 10175 CGGGACGAGACCGCGATCACCGTCGCGGACATCGACTGGGACCGCTTCTACCTCGCGTAC 10234
Qy 5821 ATCGACTGGAGCAAGATCGAACACACCTCTCAGACCAGCGACCTG------GTGAGCGCG 5874
Db 10235 TCCTCCGGTCGCCCGCAGCCCCTCGTCGAGGAGCTGCCCGAGGTGCGGCGCATCATCGAC 10294
Qy 5875 GCCCGGGAAAGGGAGCCAGCTGTCCAGCGCCCCACTCCACCGGCGG---------AGTTG 5925
Db 10295 GCACGGGACAGCGCCACGTCCGGACAGGGCGGGAGCTCCGCCCAGGGCGCCAACCCCCTG 10354
Qy 5926 CACAAAACGCTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTA 5985
Db 10355 GCCGAGCGGCTGGCCGCCGCGGCTCCCGGCGAGCGTACGGAGATCCTCCTCGGTCTCGTA 10414
Qy 5986 CGAGACCATGTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAG 6045
Db 10415 CGGGCGCAGGCCGCCGCCGTGCTCCGGATGCGTTCGCCGGAGGACGTCGCCGCCGACCGC 10474
Qy 6046 TCGTTCCGTGCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATC 6105
Db 10475 GCCTTCAAGGACATCGGCTTCGACTCGCTCGCCGGTGTCGAGCTGCGCAACAGGCTGACC 10534
Qy 6106 AAGGCAACAGGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAA 6165
Db 10535 CGGGCGACCGGGCTCCAGCTGCCCGCGACGCTCGTCTTCGACCACCCGACGCCGCTGGCC 10594





```
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.536US1
;  CURRENT APPLICATION NUMBER: US/09/988,384B
;  CURRENT FILING DATE:  2001-11-19
;  PRIOR APPLICATION NUMBER: PCT/US99/14398
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: US 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 5
;   LENGTH: 37948
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-988-384B-5

  Query Match             9.6%;  Score 2951.8;  DB 9;  Length 37948;
  Best Local Similarity   54.4%;  Pred. No. 0;
  Matches 8436;  Conservative    0;  Mismatches 5722;  Indels 1361;  Gaps   69;

Qy          8 GGATGGACGGCGGGGAAGAACCCCGCCCTGCGGCAGGGGAGGTCCTCGGAGTGGCCGACG 67
              ||  |||||     ||  | | | |   | ||| | | ||    | || | |   | |
Db       4731 GGCCGGACGATCTGGCGGCAGCGCTGGCCGCGCCTGAGGGTCTGGTCCGGGGCGTGGCTT 4790

Qy         68 AGGCGGACGGCGGCGTCGTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGG 127
                | |  |||  | ||  |  ||||| || |||||||||||| |||| |||  | |||||||
Db       4791 CCGGTGTCGGGCGAGTGGCGTTCGTGTTCCCCGGGCAGGGCACGCAGTGGGCCGGCATGG 4850

Qy        128 GAAGGGAACTTCTCGACGCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCG 187
              |    ||||| || ||| |||||| || |||  ||  |   |  ||  ||||| ||||
Db       4851 GTGCCGAACTGCTGGACTCTTCCGCGGTGTTCGCGGCGGCCATGGCCGAATGCGAGGCCG 4910

Qy        188 CGTTCGCGCCCTACGTCGACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTC 247
              |  || | || ||||||||||||||| |||||   ||  | ||| |  | || |  || |
Db       4911 CACTCTCCCCGTACGTCGACTGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGC 4970

Qy        248 CCGGGCTGGACCGGGTGGACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGG 307
              ||  |||||| ||||| || ||||| |||||       ||||||||||||| |||| ||||
Db       4971 CCACGCTGGAGCGGGTCGATGTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGG 5030

Qy        308 CCGCCCTCTGGCGCTCGCAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCG 367
              |   | | ||||     || |||||    |||     |||||  | || |||    | |||||
Db       5031 CTCGCGTGTGGCAGCACCACGGGGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCG 5090

Qy        368 AGATCGCGGCAGCCCACGTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGA 427
              ||||||| || ||  |||| | || | ||||  |||| |||||||||| || || ||||
Db       5091 AGATCGCCGCCGCGTACGTCGCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGA 5150

Qy        428 CGCTTTGGAGCCAGGCACAGACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCG 484
              | ||  | ||| || |     || |   ||| ||||| |  ||||   |  | ||| |||
Db       5151 CCCTGCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCG 5210

Qy        485 CCGCCACGCCGGATGAGCTCCTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGC 544
              |    |  ||| |   |||||   ||| | ||   ||    ||
Db       5211 CGCTGAGCGAGGACGCCGTCCTGGAGCGACTGGCCGGGTTCGACG--------------- 5255

Qy        545 GGCTCGCCGTCGCAGCCGTCAACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGG 604
              ||||  ||||||| || || ||||| |||      ||| || || ||||||| || ||
Db       5256 GGCTGTCCGTCGCCGCTGTGAACGGGCCCACCGCCACCGTGGTCTCCGGTGACCCCGTAC 5315

Qy        605 CCGTCGCGGACCTGGTGGCCGACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGG 664
                |||  || || |     |      ||||    ||| ||  ||||  | || || |
Db       5316 AGATCGAAGAGCTTGCTCGGGCGTGTGAGGCCGATGGGGTCCGTGCGCGGGTCATTCCCG 5375

Qy        665 TGGACGTTCCCGCCCACTCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCC 724
              | |||    |  |||||  ||  | | |   |||||||      | ||| |   | ||
Db       5376 TCGACTACGCGTCCCACAGCCGGCAGGTCGAGATCATCGAGAGCGAGCTCGCCGAGGTCC 5435

Qy        725 TGCTGCCCATCACCCCACGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCC 784
              |        ||| ||| |    |  | ||| | || |||  |||  |  |    |||| |
Db       5436 TCGCCGGGCTCAGCCCGCAGGCTCCGCGCGTGCCGTTCTTCTCGACACTCGAAGGCGCCT 5495

Qy        785 GCCTCGACACCCGCGAGCTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCC 844
              |  ||  |   | || ||| |||| |  ||||||||||||||| ||          ||

Db       5496 GGATCACCGAGCCCGTGCTCGACGGCGGCTACTGGTACCGCAACCTGCGCCATCGTGTGG 5555

Qy         845 GGTTCGAGCCCGCCGCCCGG---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGA 901
               | ||||  || |||| |  |   |||      |  ||||  ||    |||||||||
Db       5556 GCTTCGCCCCGGCCGTCGAGACCCTGGCCACCGACGAGGGCTTCACCCACTTCGTCGAGG 5615

Qy         902 TGAGCCCGCACCCGGTGCTGACCATGGGCCTCCAGGAG---------------------- 939
               | ||| | ||||| || || ||||||| |||||  | |
Db       5616 TCAGCGCCCACCCCGTCCTCACCATGGCCCTCCCCGGGACCGTCACCGGTCTGGCGACCC 5675

Qy         940 ----------------------------CTCGCCCCGGACCTGGGCGACACCACCGGCA 970
                                           || | | |  ||| | ||| |    | ||
Db       5676 TGCGTCGCGACAACGGCGGTCAGGACCGCCTAGTCGCCTCCCTCGCCGAAGCATGGGCCA 5735

Qy         971 CCGCCGACACCGTGATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCC 1030
               ||    | | ||   |  | || |||| | || |  |  |  |  ||  ||  |  ||
Db       5736 ACGGACTCGCGGTCGACTGGAGCCGCTCCTCCCCTCCGCGACCGGCCACCACTCCGACC 5795

Qy        1031 TGACGTCTCTCGC---------------CCAACTACGGGGGCATGGTGAGACGTCGGCGA 1075
               |  |  |  |||                |||  | |||  |    ||| ||  ||
Db       5796 TCCCCACCTACGCGTTCCAGACCGAGCGCCACTGGCTGGGCGAGATCGAGGCGCTCGCCC 5855

Qy        1076 CCACCGTCCTCTCGGCACGCCTGACCGCGCTGTCCCCCACGCAG---------------- 1119
               |  | | |    ||||    | | ||||  |   || |||| ||
Db       5856 CGGCGGGCGAGCCGGCGGTGCAGCCCGCCGTCCTCCGCACGGAGGCGGCCGAGCCGGCGG 5915

Qy        1120 -----CAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGCGGTGCTGAACG 1174
                    | | |    ||  ||| | || | |  | | |   | |||| |  | |
Db       5916 AGCTCGACCGGGACGAGCAGCTGCGCGTGATCCTGGACAAGGTCCGGGCGCAGACGGCCC 5975

Qy        1175 ACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTTCGCCCACCTCG 1234
               | |             | | |    | |  ||             || |||    |  ||
Db       5976 AGGTGCTGGGGTACGCGACAGGCGGGCAGATCGAGGTCGACCGGACCTTCCGTGAGGCCG 6035

Qy        1235 GCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGC 1294
               | | |  |||| | |  || || || ||||||||||  ||| |   |||   ||| | |
Db       6036 GTTGCACCTCCCTGACCGGCGTGGACCTGCGCAACCGGATCAACGCCGCCTTCGGCGTAC 6095

Qy        1295 GGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGCCGCGCGCCTTC 1354
               || || |   | |  | |||||||||  | |||| || |  ||  |||| | ||  || |
Db       6096 GGATGGCGCCGTCCATGATCTTCGACTTCCCCACCCCCGAGGCTCTCGCGGAGCAGCTGC 6155

Qy        1355 GGACCGCGGCGCTCGGCCACCTCGACGAGGACACCGCGCCCGTACCGGACTCACCCAGCG 1414
                  ||  | | ||  |||   | | | |  |||| | | |  | | || ||
Db       6156 TCCTCGTCGTGCACGG-------GGAGGCGGCGGCGAACCCGGCCGGTGCGGAGCCGGCT 6208

Qy        1415 GCCACGGAGGC-ACGGCAGCGGCGGACGACCCGATCGCCATCATCGGCATGGCATGCCGT 1473
               |   || |||  | |  || |  ||||| ||| | ||  ||| |||||||||| |||||
Db       6209 CCGGTGGCGGCGGCCGGTGCCGTCGACGAGCCGGTGGCGATCGTCGGCATGGCCTGCCGC 6268

Qy        1474 TTCCCGGGCGGAGTCCGGTCCCCGAAGGACCTGTGGGAGCTGGCCGCCTCGGGCGGAGAC 1533
               | || || || |||   || ||| ||||||||||  |||||  |||   ||||| |||
Db       6269 CTGCCCGGTGGGGTCGCCTCGCCGGAGGACCTGTGGCGGCTGGTGGCCGGCGGCGGGGAC 6328

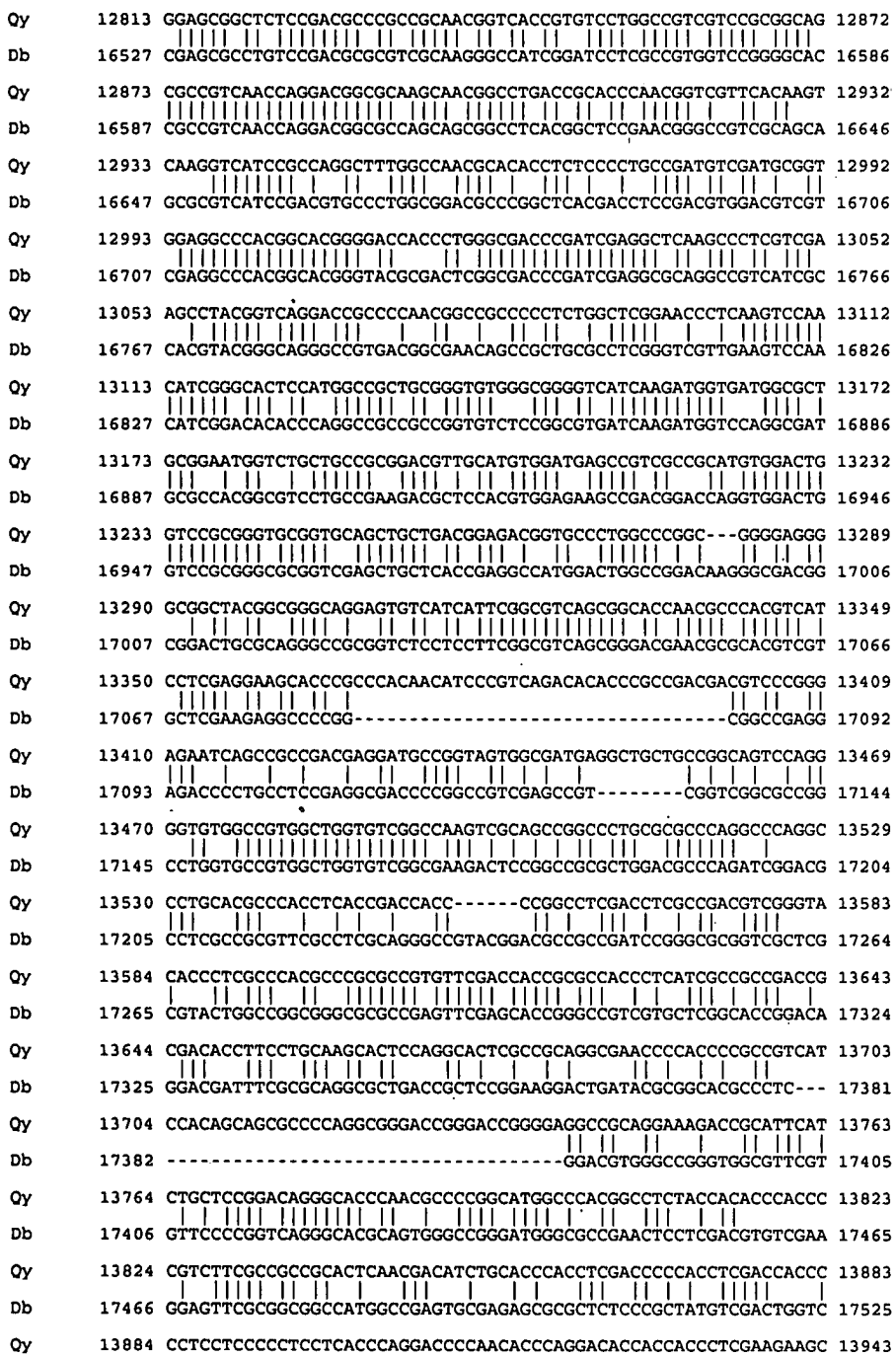
Qy        1534 GCCATCGGGCCGTTCCCCACCGACCGCGGATGGCCCACGGAACAGCGTCACGCCCAGGAC 1593
               || |||  | ||||||   |||||||| ||| |  |||   ||  ||  || |||
Db       6329 GCGATCTCGGAGTTCCCGCAGGACCGCGGCTGGGACGTGGAGGGGCTGTACCACCCGGAT 6388

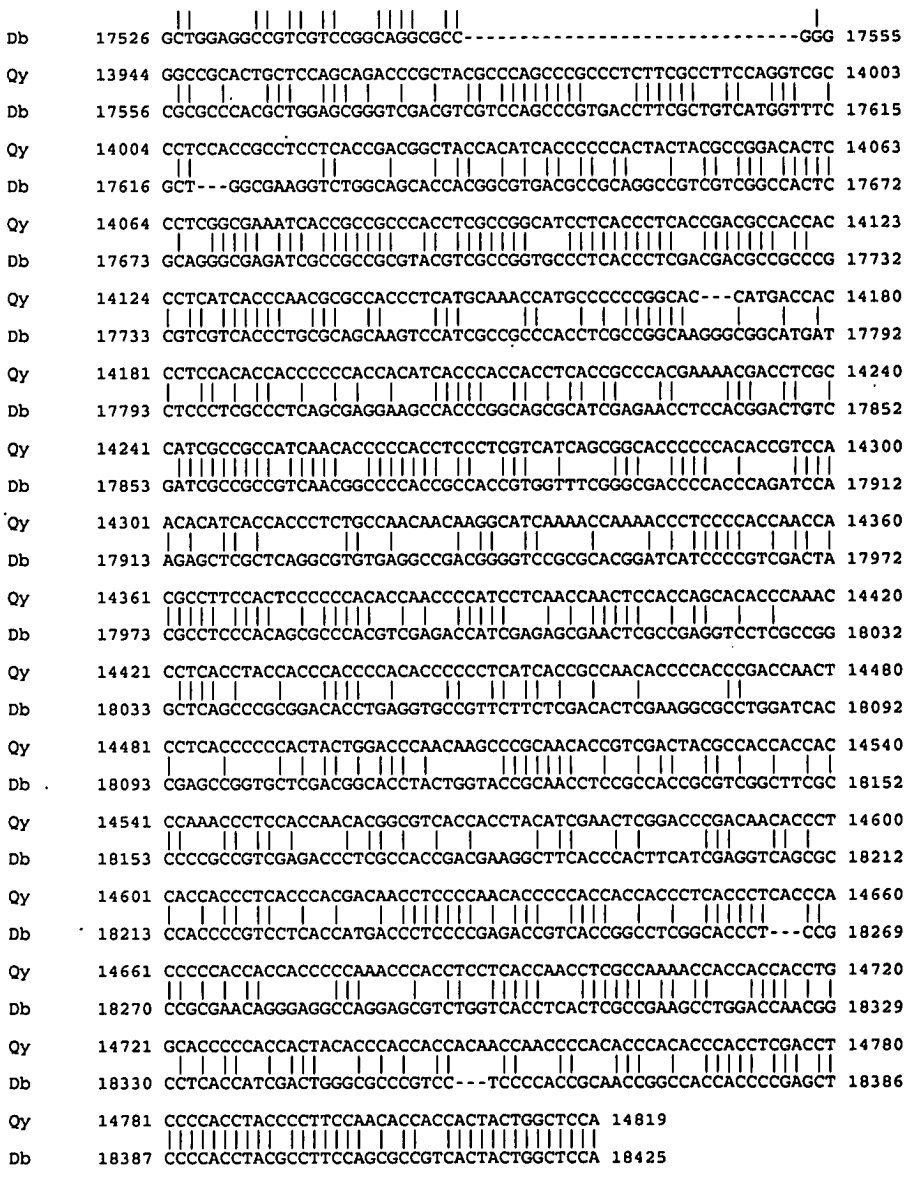
Qy        1594 CCCACGCAGCCCGGCACGTTCTATCCGCAGGGAGGCGGGTTCCTTCACGACGCGGCGCAC 1653
               ||   ||| ||||||||||  ||    |  |||||  ||| |  |  |||  ||   |
Db       6389 CCGGAGCACCCCGGCACGTCGTACGTCCGCCAGGGCGGTTTCATCGAGAACGTCGCCGGC 6448

Qy        1654 TTCGACGCCGGCTTCTTCGGAATCAGTCCACGTGAGGCACTGGCGATGGATCCGCAGCAG 1713
               |||||||| | |||||||||  |||    || || |||||  || || |||||  |||||||||
Db       6449 TTCGACGCGGCCTTCTTCGGGATCTCGCCGCGCGAGGCCCTCGCCATGGACCCGCAGCAG 6508

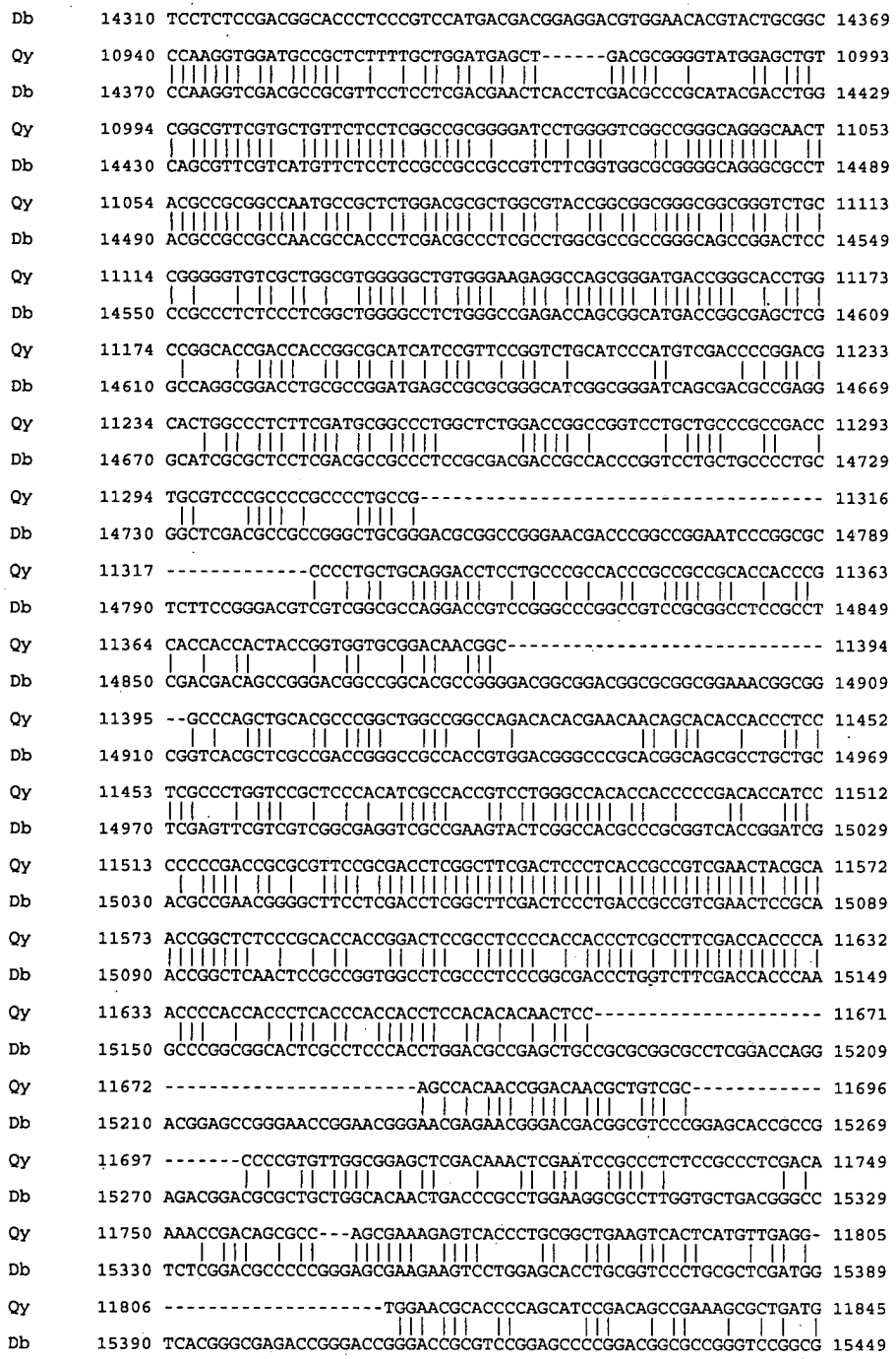
Qy        1714 CGGCTGCTGCTGGAGACGTCCTGGGAGGCGTTCGAGCGGGCGGGAATCGATCCGCTGTCG 1773
               |||||  || ||  || ||||||||||  |||||   || || |||||  |||   ||
Db       6509 CGGCTCCTCCTCGAAACCTCCTGGGAGGCCGTCGAGGACGCCGGGATCGACCCGACCTCC 6568

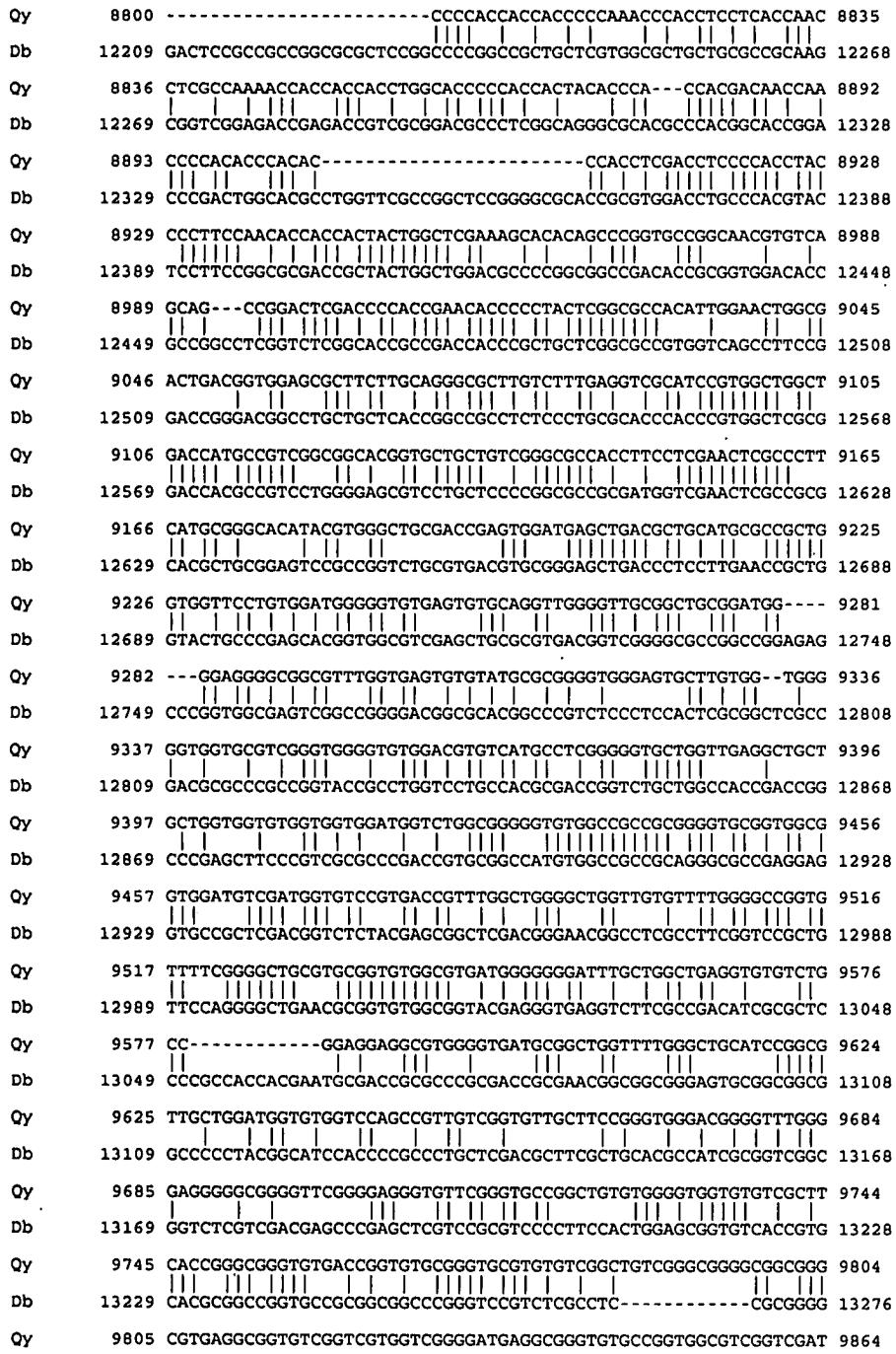
Qy        1774 GTACGCGGGTCCCGTACGGGCGTCTTCGCGGGCGCCCTCTCCTTCGACTACGGCCCGCGT 1833
               | ||  ||     |  |||||||| | ||  ||  |  ||  ||| ||||| ||| |
Db       6569 CTGCGGGGACGGCAGGTCGGCGTCTTCACTGGGGCGATGACCCACGAGTACGGGCCGAGC 6628
```

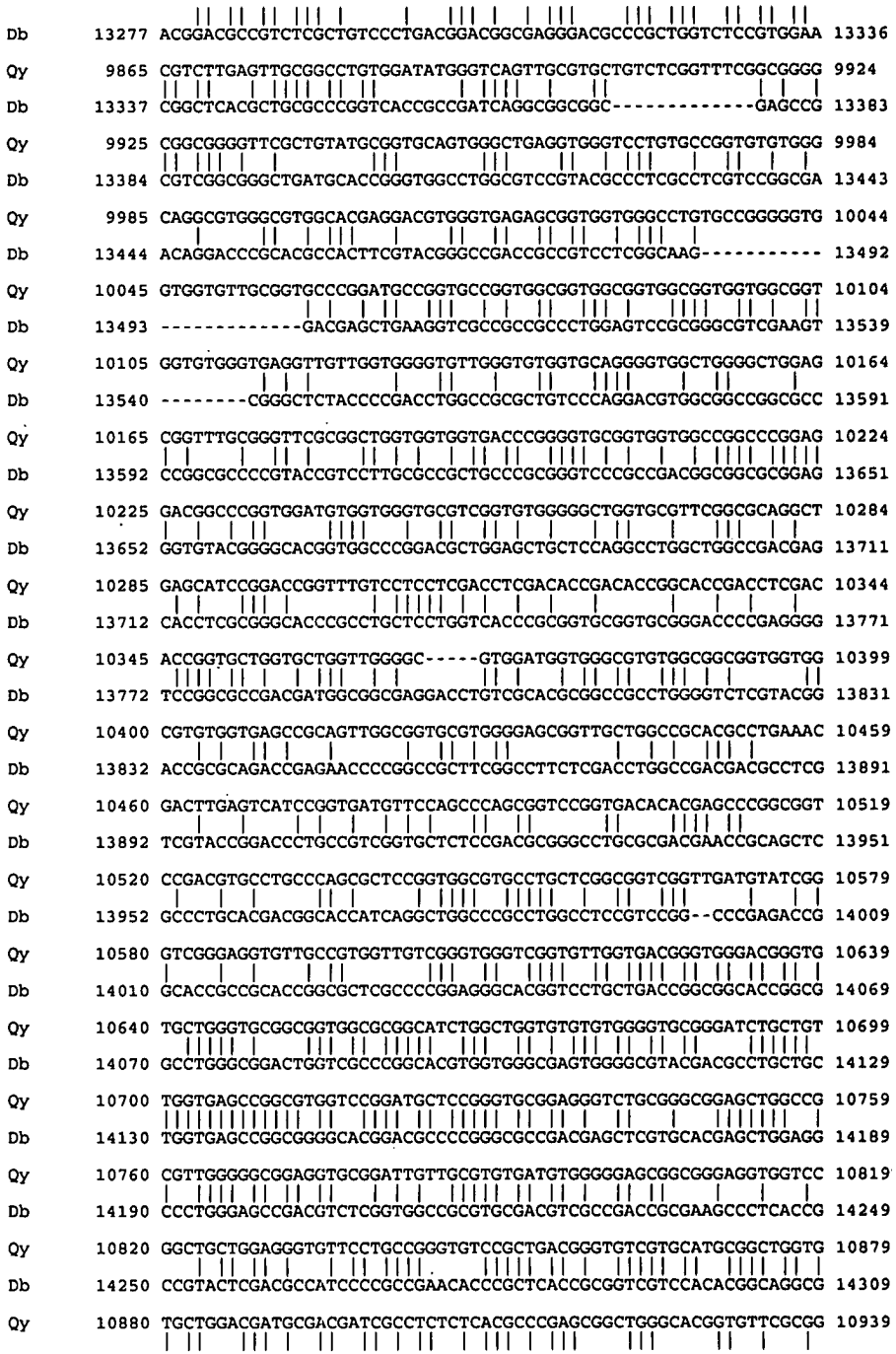


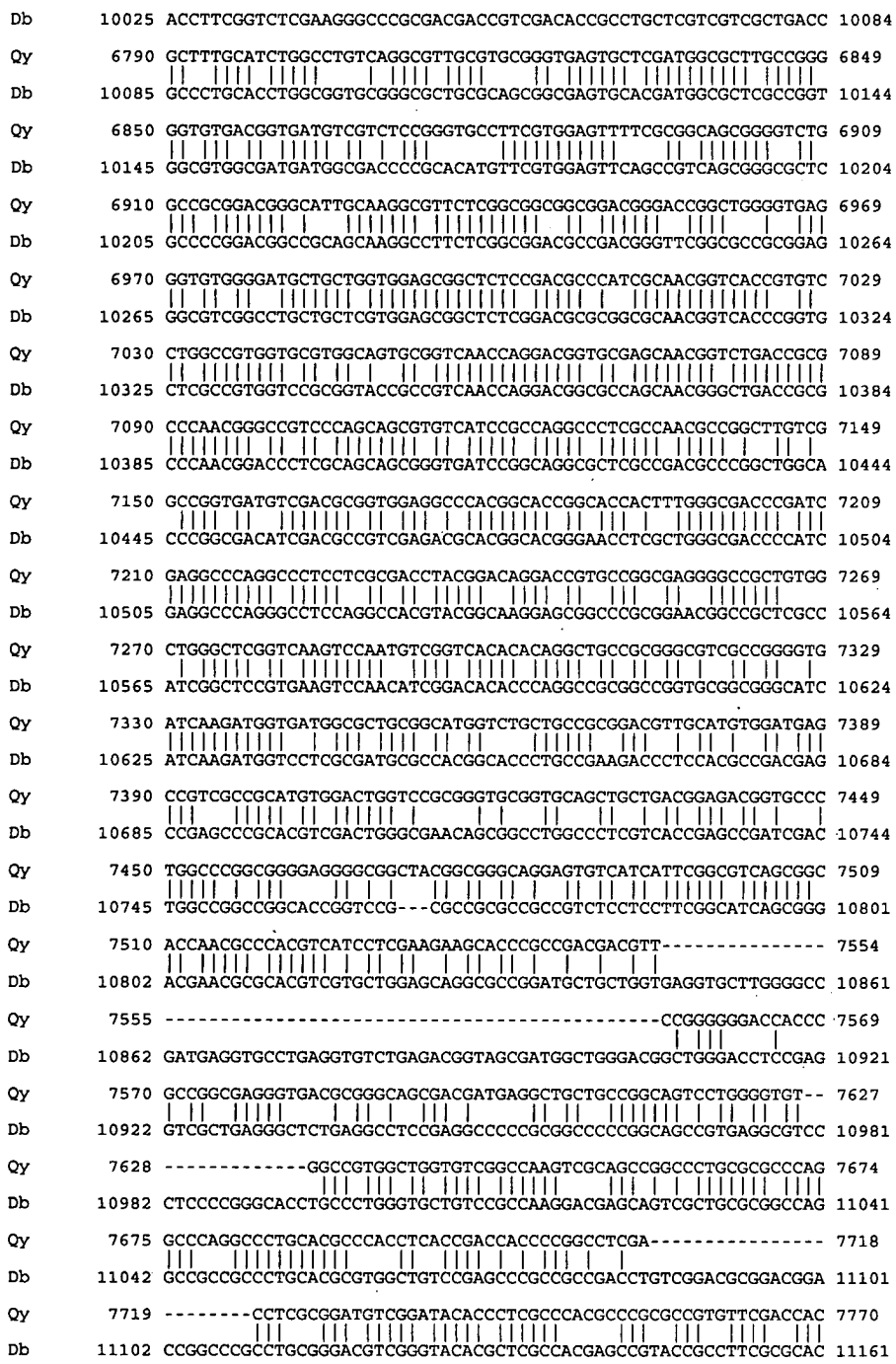


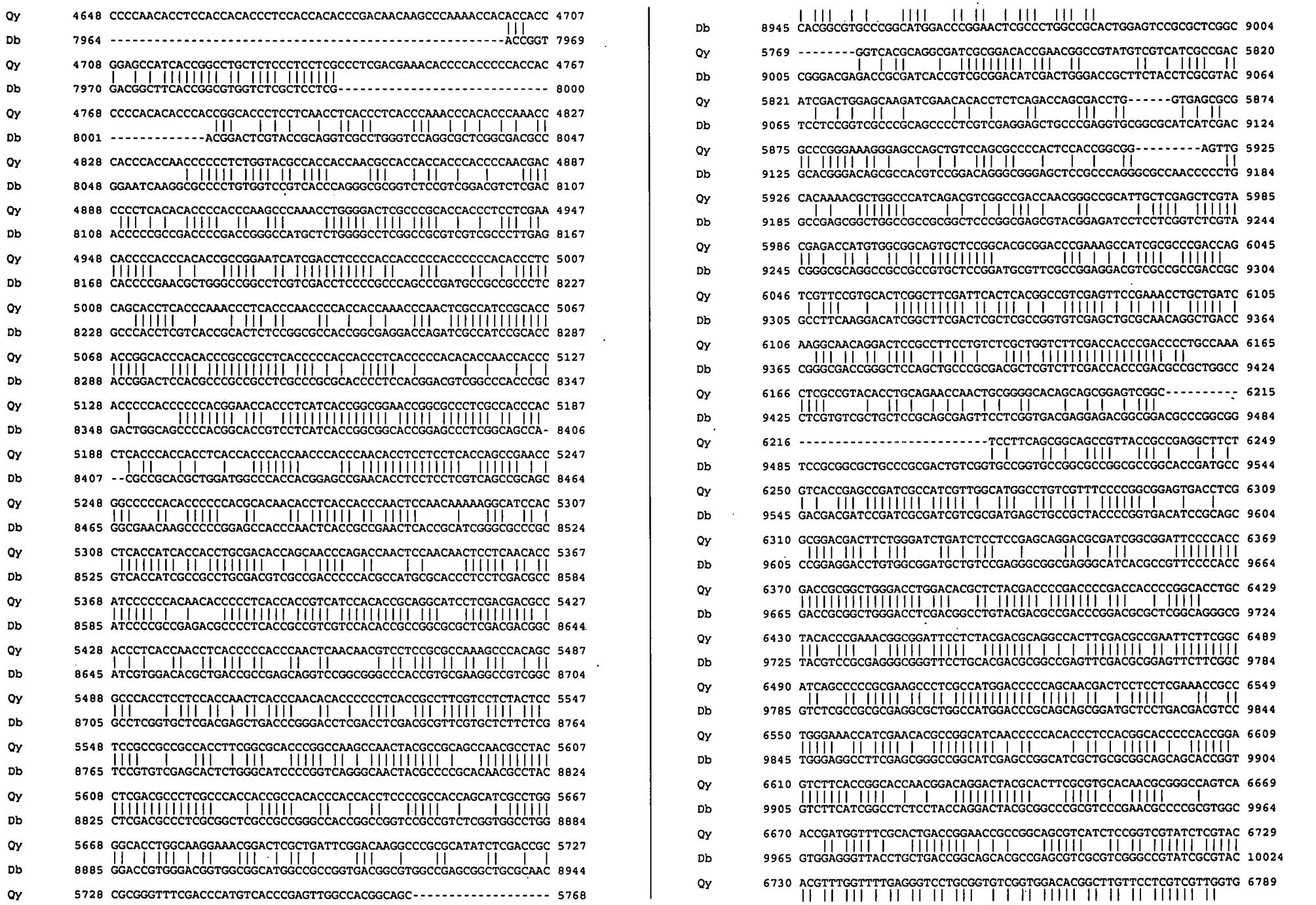
RESULT 5
US-09-988-384B-5
; Sequence 5, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.

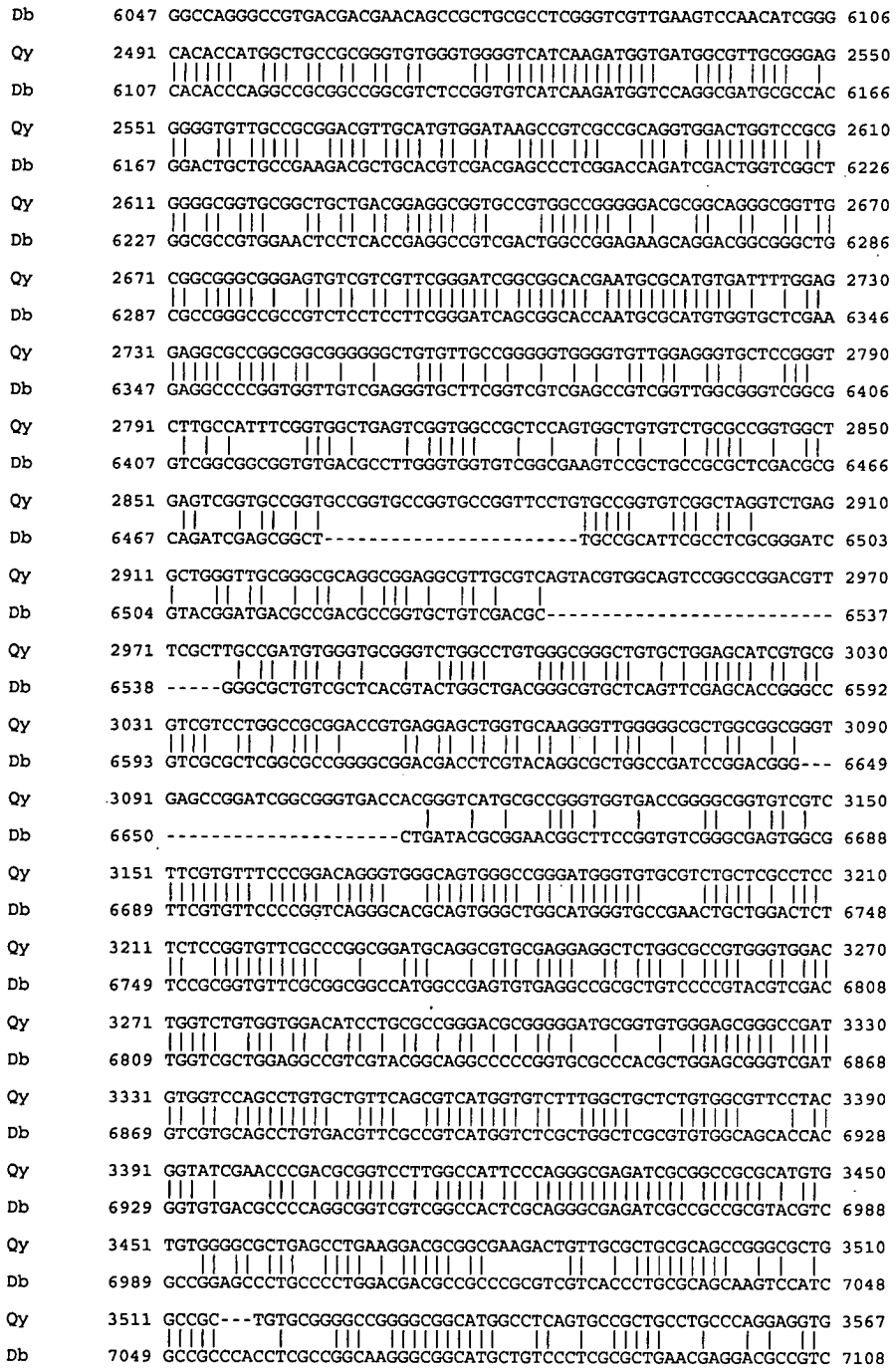

Db 14310 TCCTCTCCGACGGCACCCTCCCGTCCATGACGACGGAGGACGTGGAACACGTACTGCGGC 14369
Qy 10940 CCAAGGTGGATGCCGCTCTTTTGCTGGATGAGCT------GACGCGGGGTATGGAGCTGT 10993
Db 14370 CCAAGGTCGACGCCGCGTTCCTCCTCGACGAACTCACCTCGACGCCCGCATACGACCTGG 14429
Qy 10994 CGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATCCTGGGGTCGGCCGGGCAGGGCAACT 11053
Db 14430 CAGCGTTCGTCATGTTCTCCTCCGCCGCCGCCGTCTTCGGTGGCGCGGGGCAGGGCGCCT 14489
Qy 11054 ACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCGTACCGGCGGCGGGCGGCGGGTCTGC 11113
Db 14490 ACGCCGCCGCCAACGCCACCCTCGACGCCCTCGCCTGGCGCCGCCGGGCAGCCGGACTCC 14549
Qy 11114 CGGGGGTGTCGCTGGCGTGGGGGCTGTGGGAAGAGGCCAGCGGGATGACCGGGCACCTGG 11173
Db 14550 CCGCCCTCTCCCTCGGCTGGGGCCTCTGGGCCGAGACCAGCGGCATGACCGGCGAGCTCG 14609
Qy 11174 CCGGCACCGACCACCGGCGCATCATCCGTTCCGGTCTGCATCCCATGTCGACCCCGGACG 11233
Db 14610 GCCAGGCGGACCTGCGCCGGATGAGCCGCGCGGGCATCGGCGGGATCAGCGACGCCGAGG 14669
Qy 11234 CACTGGCCCTCTTCGATGCGGCCCTGGCTCTGGACCGGCCGGTCCTGCTGCCCGCCGACC 11293
Db 14670 GCATCGCGCTCCTCGACGCCGCCCTCCGCGACGACCGCCACCCGGTCCTGCTGCCCCTGC 14729
Qy 11294 TGCGTCCCGCCCCGCCCCTGCCG---------- 11316
Db 14730 GGCTCGACGCCGCCGGGCTGCGGGACGCGGCCGGGAACGACCCGGCCGGAATCCCGGCGC 14789
Qy 11317 ----------CCCCTGCTGCAGGACCTCCTGCCCGCCACCCGCCGCCGCACCACCCG 11363
Db 14790 TCTTCCGGGACGTCGTCGGCGCCAGGACCGTCCGGGCCCGGCCGTCCGCGGCCTCCGCCT 14849
Qy 11364 CACCACCACTACCGGTGGTGCGGACAACGGC---------- 11394
Db 14850 CGACGACAGCCGGGACGGCCGGCACGCCGGGGACGGCGGACGGCGCGGCGGAAACGGCGG 14909
Qy 11395 --GCCCAGCTGCACGCCCGGCTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCC 11452
Db 14910 CGGTCACGCTCGCCGACCGGGCCGCCACCGTGGACGGGCCCGCACGGCAGCGCCTGCTGC 14969
Qy 11453 TCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCC 11512
Db 14970 TCGAGTTCGTCGTCGGCGAGGTCGCCGAAGTACTCGGCCACGCCCGCGGTCACCGGATCG 15029
Qy 11513 CCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCA 11572
Db 15030 ACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGACCGCCGTCGAACTCCGCA 15089
Qy 11573 ACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCA 11632
Db 15090 ACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGCGACCCTGGTCTTCGACCACCCAA 15149
Qy 11633 ACCCCACCACCCTCACCCACCACCTCCACACACAACTCC---------- 11671
Db 15150 GCCCGGCGGCACTCGCCTCCCACCTGGACGCCGAGCTGCCGCGCGGCGCCTCGGACCAGG 15209
Qy 11672 ----------AGCCACAACCGGACAACGCTGTCGC---------- 11696
Db 15210 ACGGAGCCGGGAACCGGAACGGGAACGAGAACGGGACGACGGCGTCCCGGAGCACCGCCG 15269
Qy 11697 -------CCCCGTGTTGGCGGAGCTCGACAAACTCGAATCCGCCCTCTCCGCCCTCGACA 11749
Db 15270 AGACGGACGCGCTGCTGGCACAACTGACCCGCCTGGAAGGCGCCTTGGTGCTGACGGGCC 15329
Qy 11750 AAACCGACAGCGCC---AGCGAAAGAGTCACCCTGCGGCTGAAGTCACTCATGTTGAGG- 11805
Db 15330 TCTCGGACGCCCCCGGGAGCGAAGAAGTCCTGGAGCACCTGCGGTCCCTGCGCTCGATGG 15389
Qy 11806 ----------TGGAACGCACCCCAGCATCCGACAGCCGAAAGCGCTGATG 11845
Db 15390 TCACGGGCGAGACCGGGACCGGGACCGCGTCCGGAGCCCCGGACGGCGCCGGGTCCGGCG 15449
Qy 11846 ACGACGA---------- 11852
Db 15450 CCGAGGACCGGCCCTGGGCGGCCGGGGACGGAGCCGGGGGCGGGAGTGAGGACGGCGCGG 15509
Qy 11853 -------GAAGTTCACATCGGCAACAGAGGCTGAGATTTTCAAATTCATTGACAACGACC 11905
Db 15510 GAGTGCCGGACTTCATGAACGCCTCGGCCGAGGAACTCTTCGGCCTCCTCGACCAGGACC 15569
Qy 11906 TCGGCCTGTCCTGAACCGGACGCCTGCCACTCCGCCCGTATCCG---------- 11949
Db 15570 CCAGCACGGACTGATCCCTGCCGCACGGTCGCCTCCCGCCCCGGACCCCGTCCCGGGCAC 15629
Qy 11950 ----------CTGGGCCCTGCTAGGACGTGAAT 11972
Db 15630 CTCGACTCGAATCACTTCATGCGCGCCTCGGGCGCCTCCAGGAACTCAAGGGGACAGCGT 15689
Qy 11973 GCAATTGGCGAATGAAGCGAAGCTCCTGGAATACCTCAAGCGCGTCACTGCGGACCTGGA 12032
Db 15690 GTCCACGGTGAACGAAGAGAAGTACCTCGACTACCTGCGTCGTGCCACGGCGGACCTCCA 15749
Qy 12033 CCGCACTCGCCGTCGCCTGTACGAGGTGGTCGAGCGTGAGCAGGAGCCGATCGCGATTGT 12092
Db 15750 CGAGGCCCGTGGCCGCCTCCGCGAGCTGGAGGCGAAGGCGGGCGAGCCGGTGGCGATCGT 15809
Qy 12093 GGGGATGGCGTGTCGTTACCCAGGCGGGGCGACGTCACCCACGCGACTGTGGCATCTCGT 12152
Db 15810 CGGCATGGCCTGCCGCCTGCCCGGCGGCGTCGCCTCGCCCGAGGACCTGTGGCGGCTGGT 15869
Qy 12153 CAAGTCCCAGACGGACGCTATCGGGGAGTTCCCGACCGACCGTGGATGGAACCTGGAGCA 12212
Db 15870 GGCCGGCGGCGAGGACGCGATCTCGGAGTTCCCCCAGGACCGCGGCTGGGACGTGGAGGG 15929
Qy 12213 GCTCTACGACCCGGACCCCGACCGCTCAGGAACCAGTTACACGCGCAGCGGAGGGTTTCT 12272
Db 15930 CCTGTACGACCCGAACCCGGAGGCCACGGGCAAGAGTTACGCCCGCGAGGCCGGATTCCT 15989
Qy 12273 CTATGACGCGGGCGACTTCGACGCCGCGTTCTTCGAGTTGTCACCGCGTGAGGCGCTGGC 12332
Db 15990 GTACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCGCCGCGCGAGGCCCTCGC 16049
Qy 12333 AATGGACCCGCAGCAGCGCCTGCTGCTCGAAACCACTTGGGAAACGTTCGAACAGGGCGG 12392
Db 16050 CATGGACCCGCAGCAGCGTCTCCTCCTGGAGGCCTCCTGGGAGGCGTTCGAGCACGCCGG 16109
Qy 12393 AATCGACCCGAGGTCCATGCGCGGAAGCCGGACCGGGGTTTTCGTGGGGATCAATCCGGA 12452
Db 16110 GATCCCGGCGGCCACCGCGCGCGGCACCTCGGTCGGCGTCTTCACCGGCGTGATGTACCA 16169
Qy 12453 GGACTACACCACCGGATACACACATCAGCCCTCAAACGCAGTCGAGGGCTACCTGCTCAC 12512
Db 16170 CGACTACGCCACCCGTCTCACCGAT---GTCCCGGAGGGCATCGAGGGCTACCTGGGCAC 16226
Qy 12513 TGGCAGCGCGGCAAGCATTGCGTCAGGCCGTATCTCCTACAACTTCGGGCTCGAAGGCCC 12572
Db 16227 CGGCAACTCCGGCAGTGTCGCCTCGGGCCGCGTCGCGTACACGCTTGGCCTGGAGGGGCC 16286
Qy 12573 TGCGATCACTATCGACACCGCGTGTTCCTCCTCGCTCGTCGCCCTGCATCTGGCCTGCCA 12632
Db 16287 GGCCGTCACGGTCGACACCGCCTGCTCGTCCTCGCTGGTCGCCCTGCACCTCGCCGTGCA 16346
Qy 12633 AGCGCTCCGGTCCGGTGAATGCACCATGGCGCTCGCAGGCGGCGCCTCCGTCATGGCCAC 12692
Db 16347 GGCCCTGCGCAAGGGCGAGGTCGACATGGCGCTCGCCGGCGGCGTGACGGTCATGTCGAC 16406
Qy 12693 TCCCTTCGTCTTCACCGAGTTCTCTCGCCAGCGGGGCCTGGCCGCAGACGGCCGGTGCAA 12752
Db 16407 GCCCAGCACCTTCGTCGAGTTCAGCCGTCAGCGCGGGCTGGCGCCGGACGGCCGGTCGAA 16466
Qy 12753 GGCGTTTTCGGCGGCGGCGGACGGGACCGGCTGGTCCGAGGGTGTGGGGATGCTGCTGGT 12812
Db 16467 GTCCTTCTCGTCGACGGCCGACGGCACCAGCTGGTCCGAGGGCGTCGGCGTCCTCCTCGT 16526

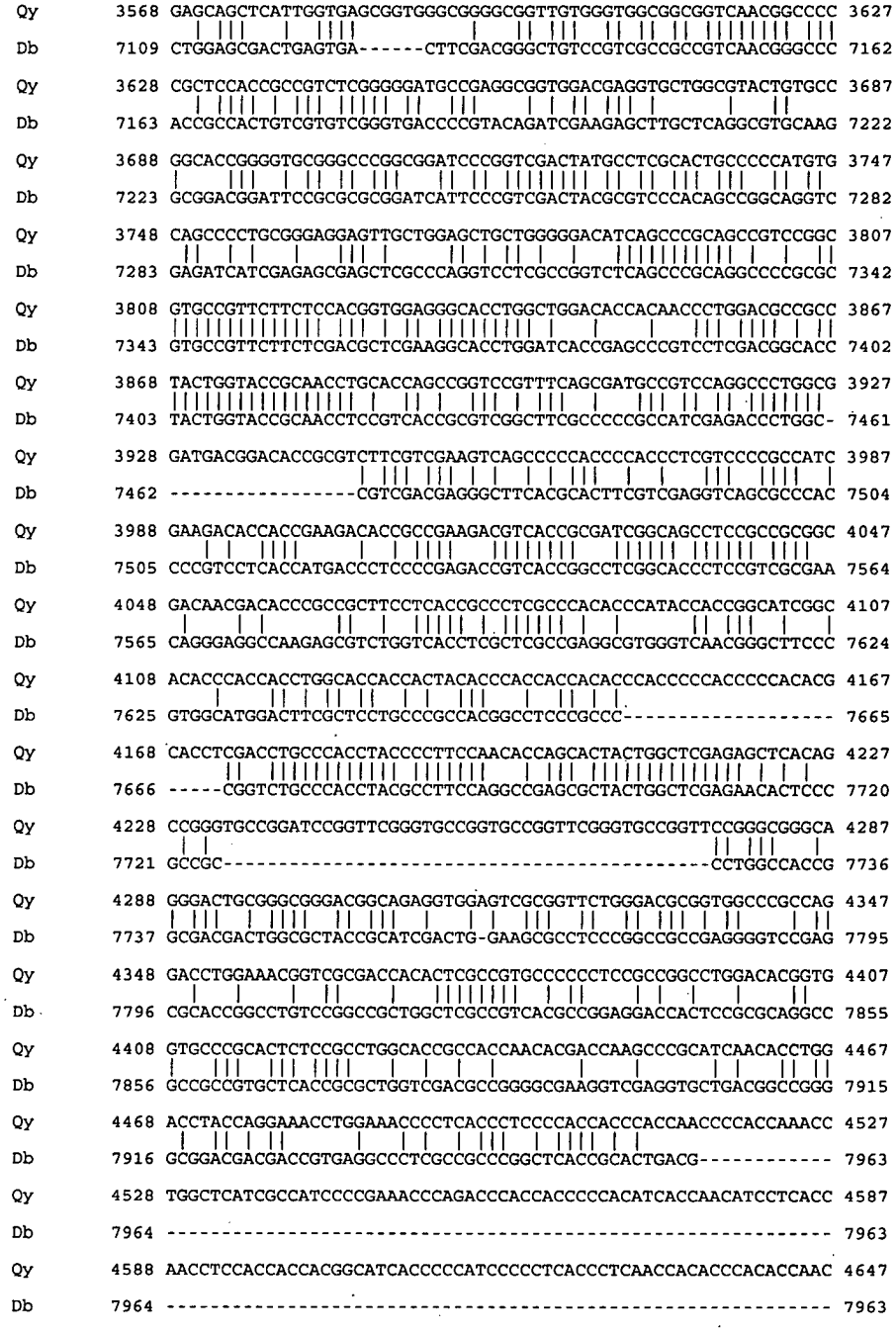


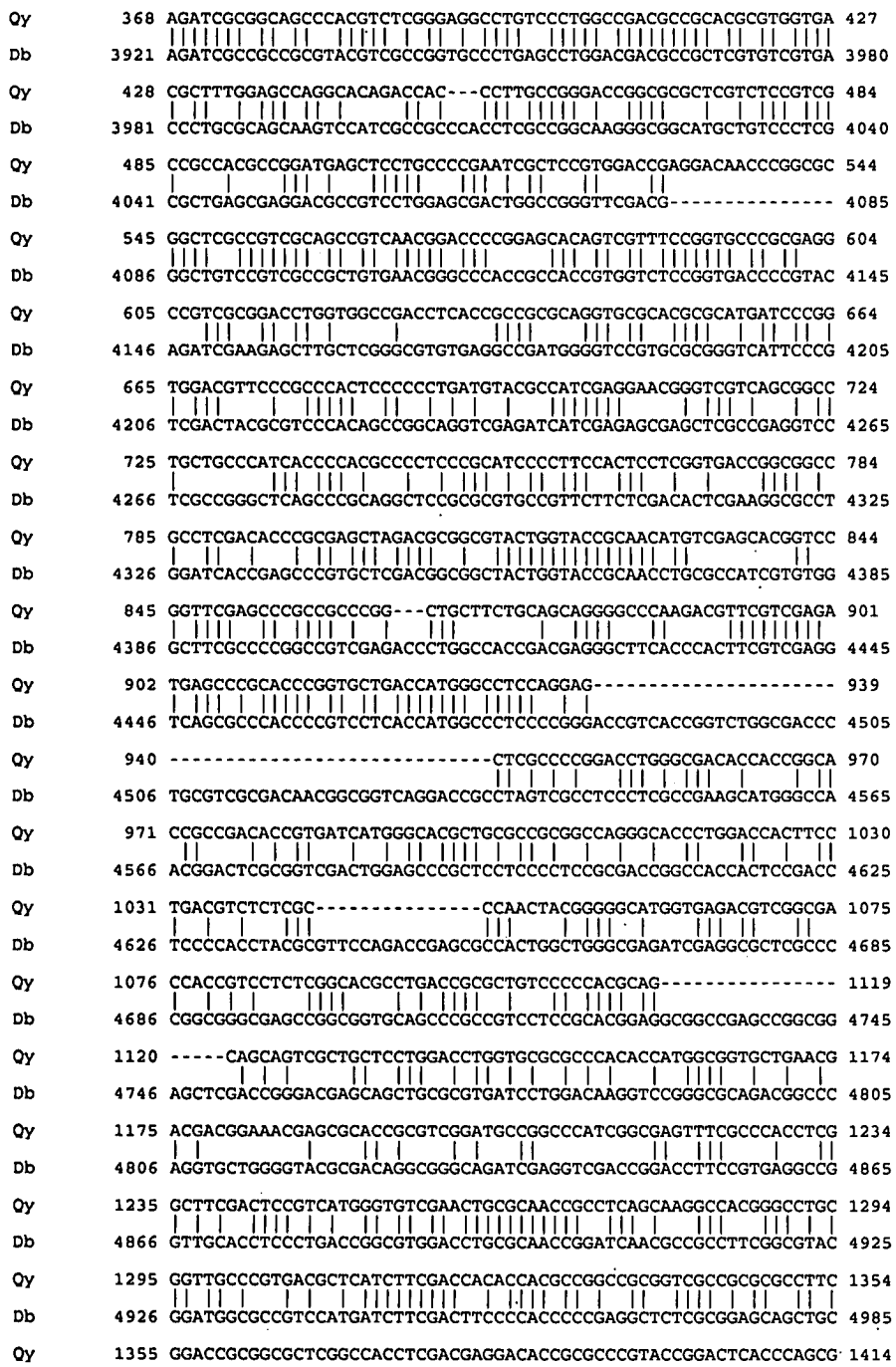


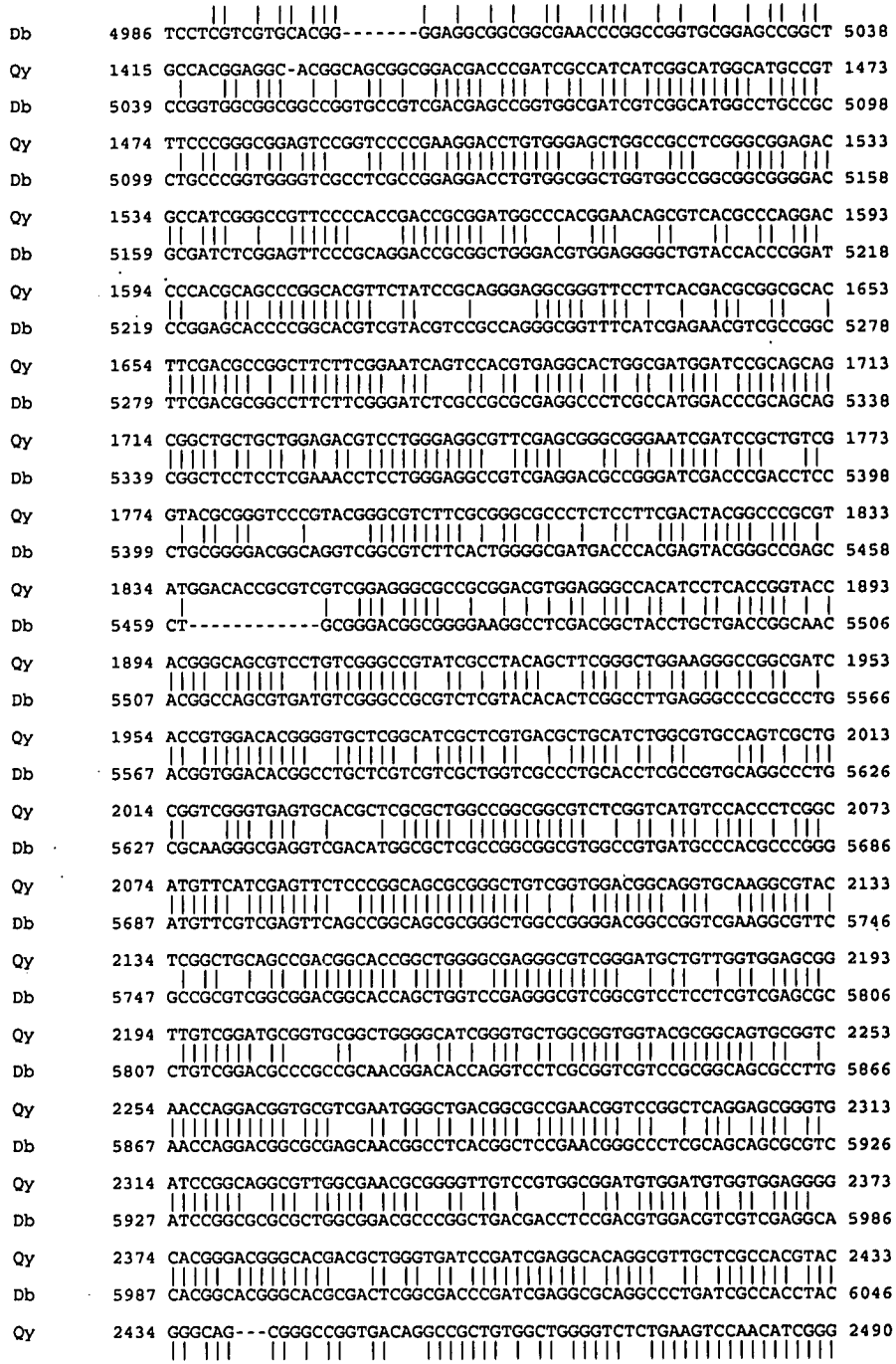

Db 10025 ACCTTCGGTCTCGAAGGGCCCGCGACGACCGTCGACACCGCCTGCTCGTCGTCGCTGACC 10084
Qy 6790 GCTTTGCATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGG 6849
Db 10085 GCCCTGCACCTGGCGGTGCGGGCGCTGCGCAGCGGCGAGTGCACGATGGCGCTCGCCGGT 10144
Qy 6850 GGTGTGACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTG 6909
Db 10145 GGCGTGGCGATGATGGCGACCCCGCACATGTTCGTGGAGTTCAGCCGTCAGCGGGCGCTC 10204
Qy 6910 GCCGCGGACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAG 6969
Db 10205 GCCCCGGACGGCCGCAGCAAGGCCTTCTCGGCGGACGCCGACGGGTTCGGCGCCGCGGAG 10264
Qy 6970 GGTGTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTC 7029
Db 10265 GGCGTCGGCCTGCTGCTCGTGGAGCGGCTCTCGGACGCGCGGCGCAACGGTCACCCGGTG 10324
Qy 7030 CTGGCCGTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCG 7089
Db 10325 CTCGCCGTGGTCCGCGGTACCGCCGTCAACCAGGACGGCGCCAGCAACGGGCTGACCGCG 10384
Qy 7090 CCCAACGGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCG 7149
Db 10385 CCCAACGGACCCTCGCAGCAGCGGGTGATCCGGCAGGCGCTCGCCGACGCCCGGCTGGCA 10444
Qy 7150 GCCGGTGATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATC 7209
Db 10445 CCCGGCGACATCGACGCCGTCGAGACGCACGGCACGGGAACCTCGCTGGGCGACCCCATC 10504
Qy 7210 GAGGCCCAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGG 7269
Db 10505 GAGGCCCAGGGCCTCCAGGCCACGTACGGCAAGGAGCGGCCCGCGGAACGGCCGCTCGCC 10564
Qy 7270 CTGGGCTCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTG 7329
Db 10565 ATCGGCTCCGTGAAGTCCAACATCGGACACACCCAGGCCGCGGCCGGTGCGGCGGGCATC 10624
Qy 7330 ATCAAGATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAG 7389
Db 10625 ATCAAGATGGTCCTCGCGATGCGCCACGGCACCCTGCCGAAGACCCTCCACGCCGACGAG 10684
Qy 7390 CCGTCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCC 7449
Db 10685 CCGAGCCCGCACGTCGACTGGGCGAACAGCGGCCTGGCCCTCGTCACCGAGCCGATCGAC 10744
Qy 7450 TGGCCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGC 7509
Db 10745 TGGCCGGCCGGCACCGGTCCG---CGCCGCGCCGCCGTCTCCTCCTTCGGCATCAGCGGG 10801
Qy 7510 ACCAACGCCCACGTCATCCTCGAAGAAGCACCCGCCGACGACGTT---------- 7554
Db 10802 ACGAACGCGCACGTCGTGCTGGAGCAGGCGCCGGATGCTGCTGGTGAGGTGCTTGGGGCC 10861
Qy 7555 ----------CCGGGGGGACCACCC 7569
Db 10862 GATGAGGTGCCTGAGGTGTCTGAGACGGTAGCGATGGCTGGGACGGCTGGGACCTCCGAG 10921
Qy 7570 GCCGGCGAGGGTGACGCGGGCAGCGACGATGAGGCTGCTGCCGGCAGTCCTGGGGTGT-- 7627
Db 10922 GTCGCTGAGGGCTCTGAGGCCTCCGAGGCCCCCGCGGCCCCCGGCAGCCGTGAGGCGTCC 10981
Qy 7628 ----------GGCCGTGGCTGGTGTCGGCCAAGTCGCAGCCGGCCCTGCGCGCCCAG 7674
Db 10982 CTCCCCGGGCACCTGCCCTGGGTGCTGTCCGCCAAGGACGAGCAGTCGCTGCGCGGCCAG 11041
Qy 7675 GCCCAGGCCCTGCACGCCCACCTCACCGACCACCCCGGCCTCGA---------- 7718
Db 11042 GCCGCCGCCCTGCACGCGTGGCTGTCCGAGCCCGCCGCCGACCTGTCGGACGCGGACGGA 11101
Qy 7719 --------CCTCGCGGATGTCGGATACACCCTCGCCCACGCCCGCGCCGTGTTCGACCAC 7770
Db 11102 CCGGCCCGCCTGCGGGACGTCGGGTACACGCTCGCCACGAGCCGTACCGCCTTCGCGCAC 11161
Qy 7771 CGCGCCACCCTCATCGCCGCGGACCGCGACACGTTCCTGCAAGCACTCCAGGCACTCGCC 7830
Db 11162 CGCGCCGCCGTGACCGCCGCCGACCGGGACGGGTTCCTGGACGGGCTGGCCACGCTGGCC 11221
Qy 7831 GCAGGCGAGCCCCACCCCGCCGTCATCCACAGCAGCGCCCCGGGCGGGACCGGGACCGGG 7890
Db 11222 CAGGGCGGCACCTCGGCCCACGTCCACCTGGACACCGCCCGGGACGGCACC--------- 11272
Qy 7891 GAGGCCGCAGGAAAGACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATG 7950
Db 11273 ----------ACCGCGTTCCTCTTCACCGGCCAGGGCAGTCAGCGCCCCGGCGCC 11317
Qy 7951 GCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACC 8010
Db 11318 GGCCGTGAGCTGTACGACCGGCACCCCGTCTTCGCCCGGGCGCTCGACGAGATCTGCGCC 11377
Qy 8011 CACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGAC 8070
Db 11378 CACCTCGACGGTCACCTCGAACTGCCCCTGCTCGACGTGAT------GTTCGCGGCCGAG 11431
Qy 8071 AACGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCC 8130
Db 11432 GGCAGCGCGGAGGCCGCGCTGCTCGACGAGACGCGGTACACGCAGTGCGCGCTGTTCGCC 11491
Qy 8131 TTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTAC 8190
Db 11492 CTGGAGGTCGCGCTCTTCCGGCTCGTC---GAGAGCTGGGGCATGCGGCCGGCCGCACTG 11548
Qy 8191 GCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACC 8250
Db 11549 CTCGGTCACTCGGTCGGCGAGATCGCCGCCGCGCACGTCGCCGGTGTGTTCTCGCTCGCC 11608
Qy 8251 GACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCC---GGC 8307
Db 11609 GACGCCGCCCGCCTGGTCGCCGCGCGCGGCCGGCTCATGCAGGAGCTGCCCGCCGGTGGC 11668
Qy 8308 ACCATGACCACCCTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAA 8367
Db 11669 GCGATGCTCGCCGTCCAGGCCGCGGAGGACGAGATCCGCGTGTGGCTGGAGACGGAGGAG 11728
Qy 8368 AACGAC----------CTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGC 8418
Db 11729 CGGTACGCGGGACGTCTGGACGTCGCCGCCGTCAACGGCCCCGAGGCCGCCGTCCTGTCC 11788
Qy 8419 GGCACCCCCCACACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACC 8478
Db 11789 GGCGACGCGGACGCGGCGCGGGAGGCGGAGGCGTACTGGTCCGGGCTCGGCCGCAGGACC 11848
Qy 8479 AAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAA 8538
Db 11849 CGCGCGCTGCGGGTCAGCCACGCCTTCCACTCCGCGCACATGGACGGCATGCTCGACGGG 11908
Qy 8539 CTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCC 8598
Db 11909 TTCCGCGCCGTCCTGGAGACGGTGGAGTTCCGGCGCCCCTCCCTGACCGTGGTCTCGAAC 11968
Qy 8599 AACA----------CCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAA 8646
Db 11969 GTCACCGGCCTGGCCGCCGGCCCGGACGACCTGTGCGACCCCGAGTACTGGGTCCGGCAC 12028
Qy 8647 GCCCGCAACACCGTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACC 8706
Db 12029 GTCCGCGGCACCGTCCGCTTCCTCGACGGCGTCCGTGTCCTGCGCGACCTCGGCGTGCGG 12088
Qy 8707 ACCTACATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCC 8766
Db 12089 ACCTGCCTGGAGCTGGGCCCCGACGGGGTCCTCACCGCCATGGCGGCCGACGGCCTCGCG 12148
Qy 8767 AACCCCCCCACCACCACCCTCACCCTCACCCAC---------- 8799
Db 12149 GACACCCCCGCGGATTCCGCTGCCGGCTCCCCCGTCGGCTCTCCCGCCGGCTCTCCCGCC 12208

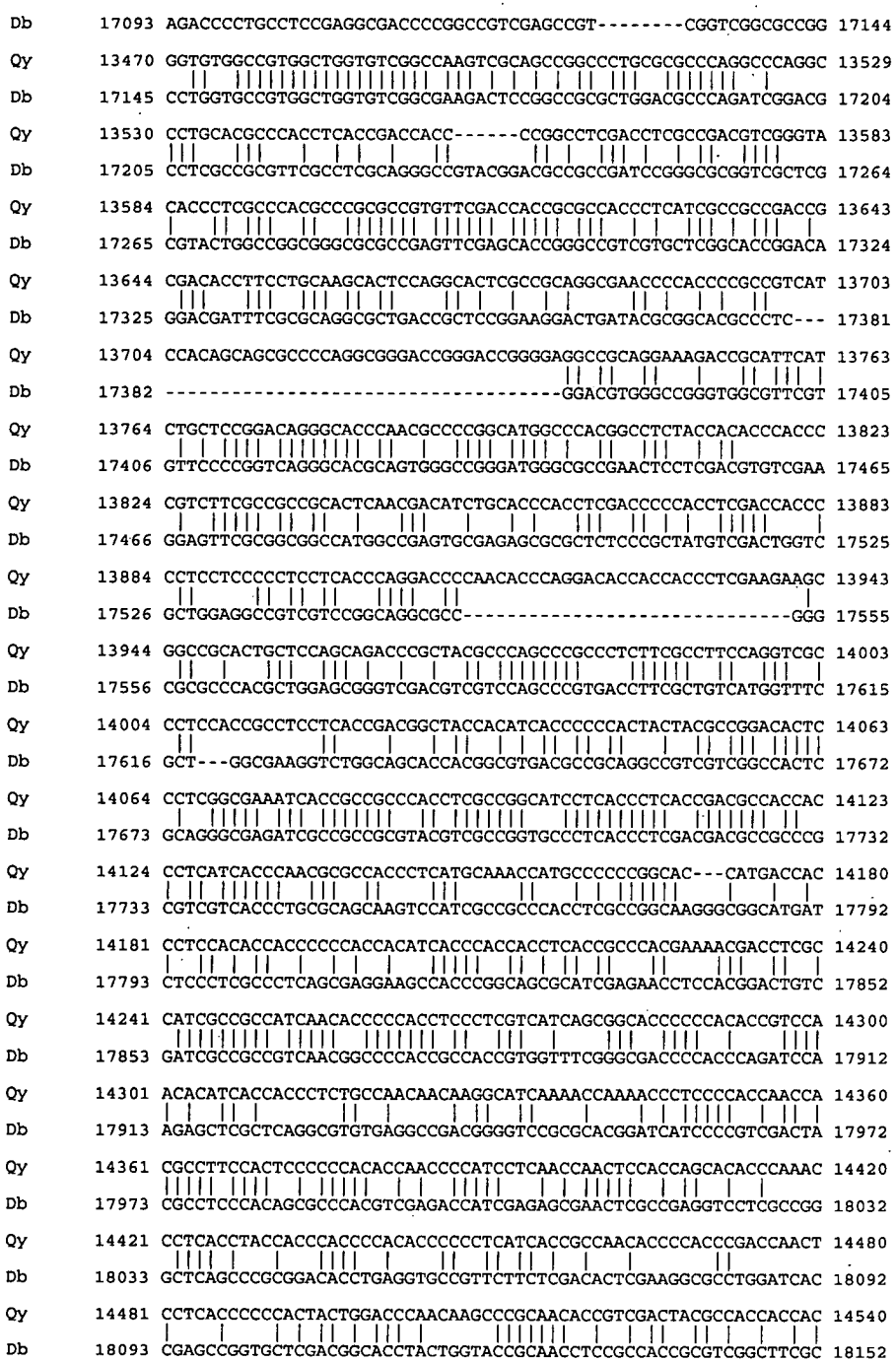


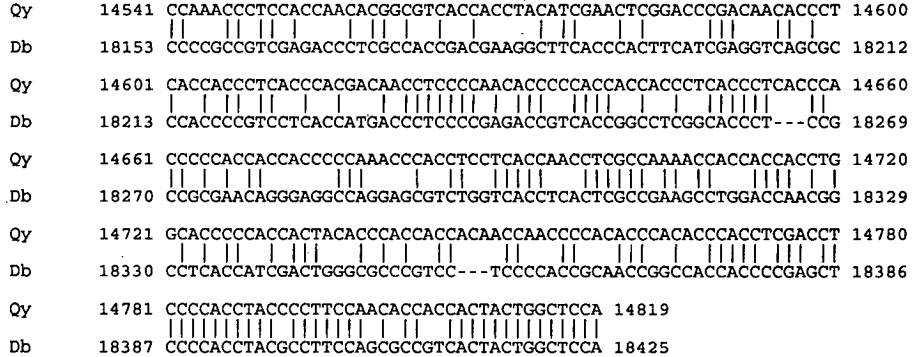






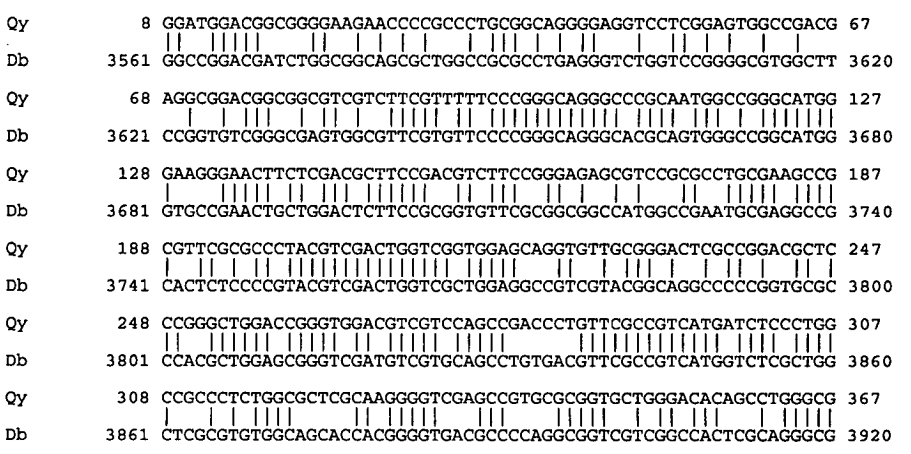


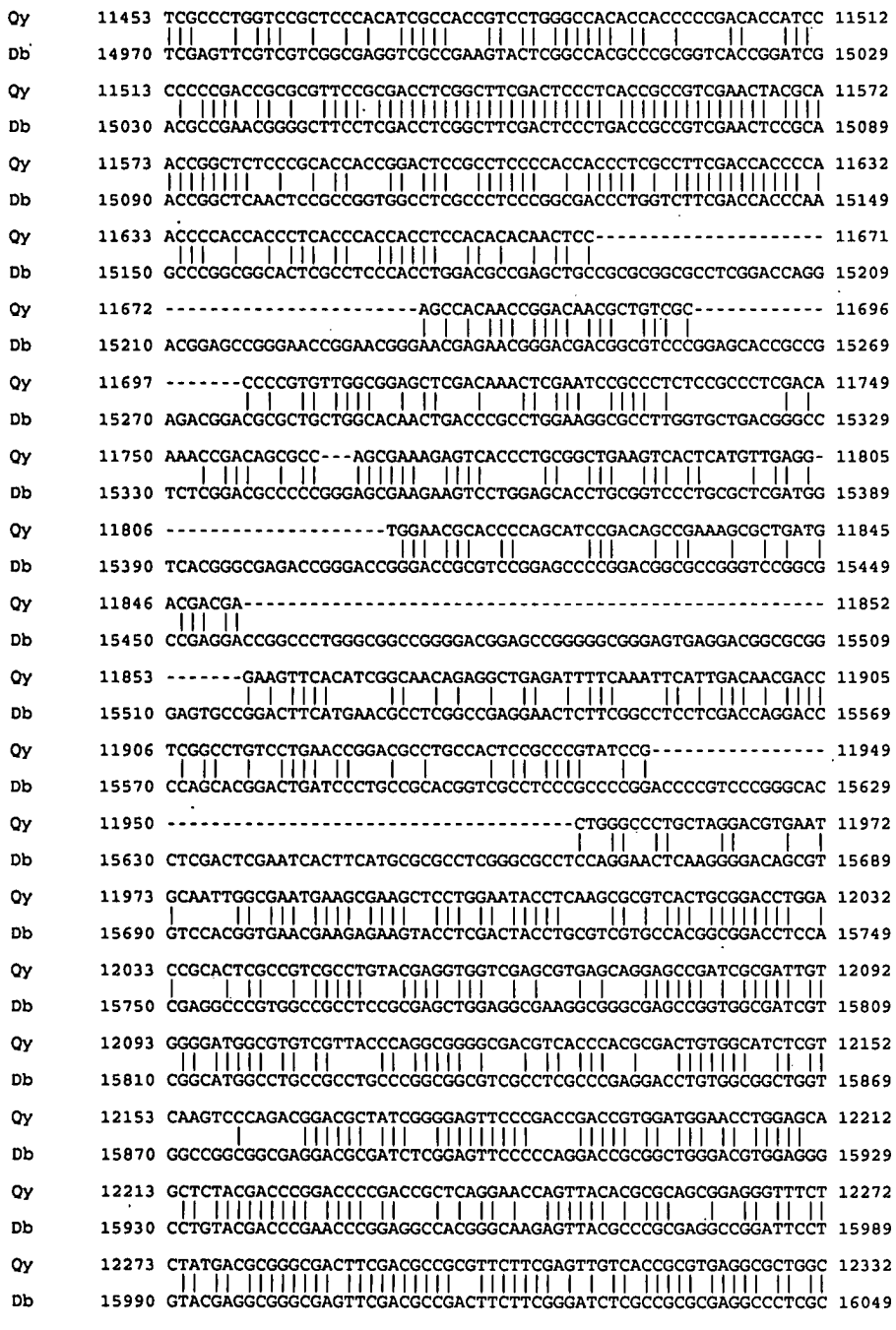


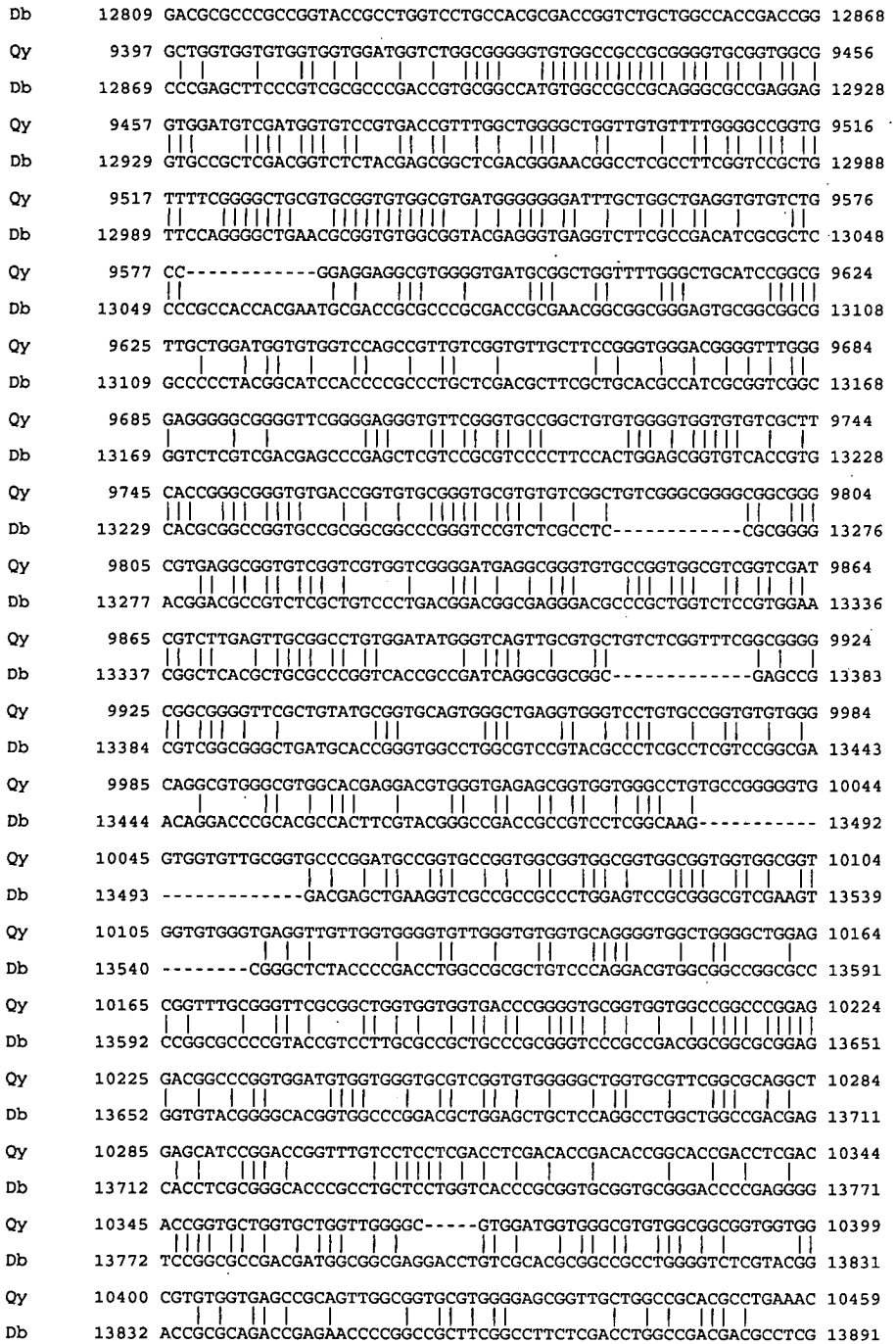


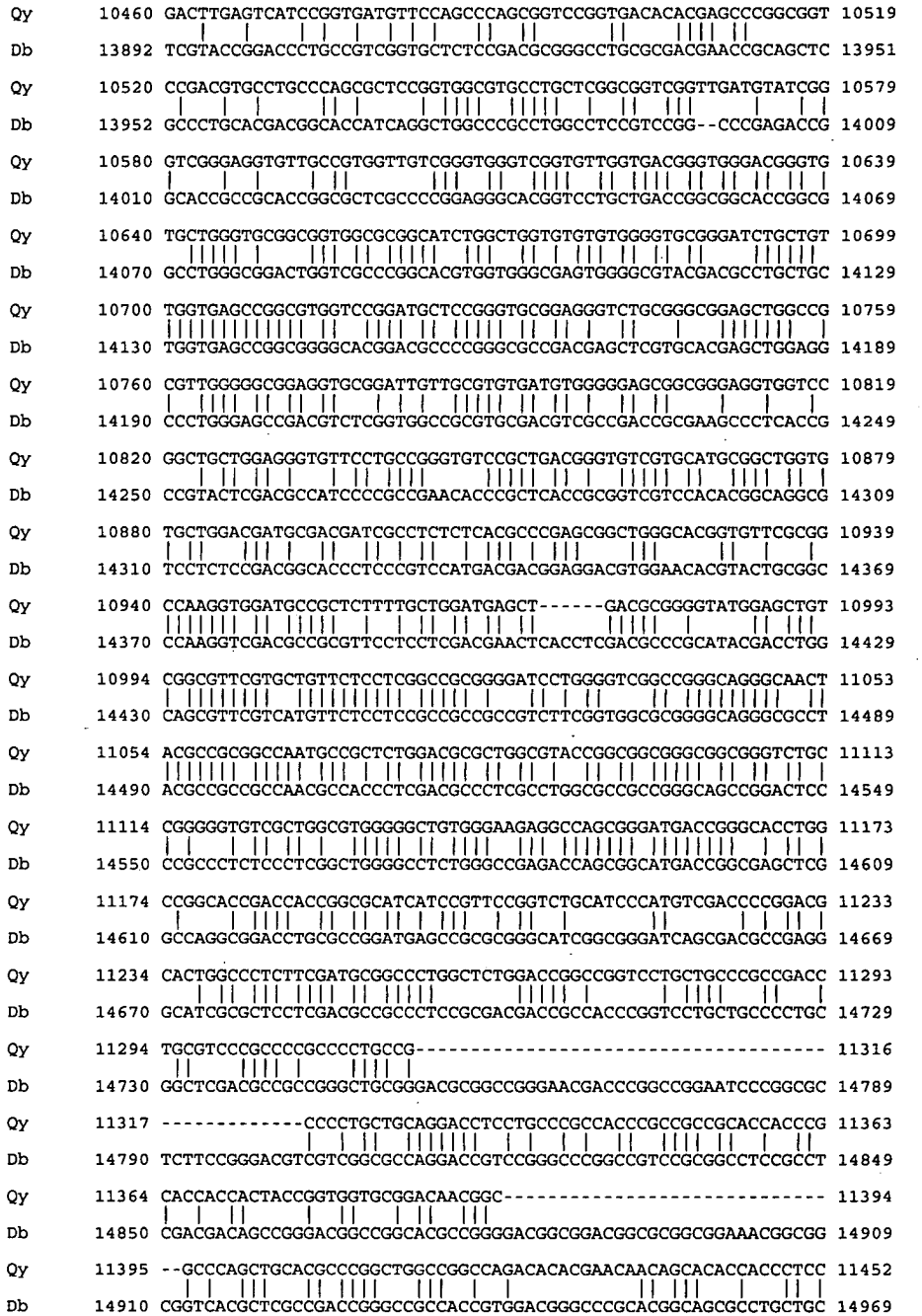
RESULT 4
US-09-861-289-5
; Sequence 5, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-5

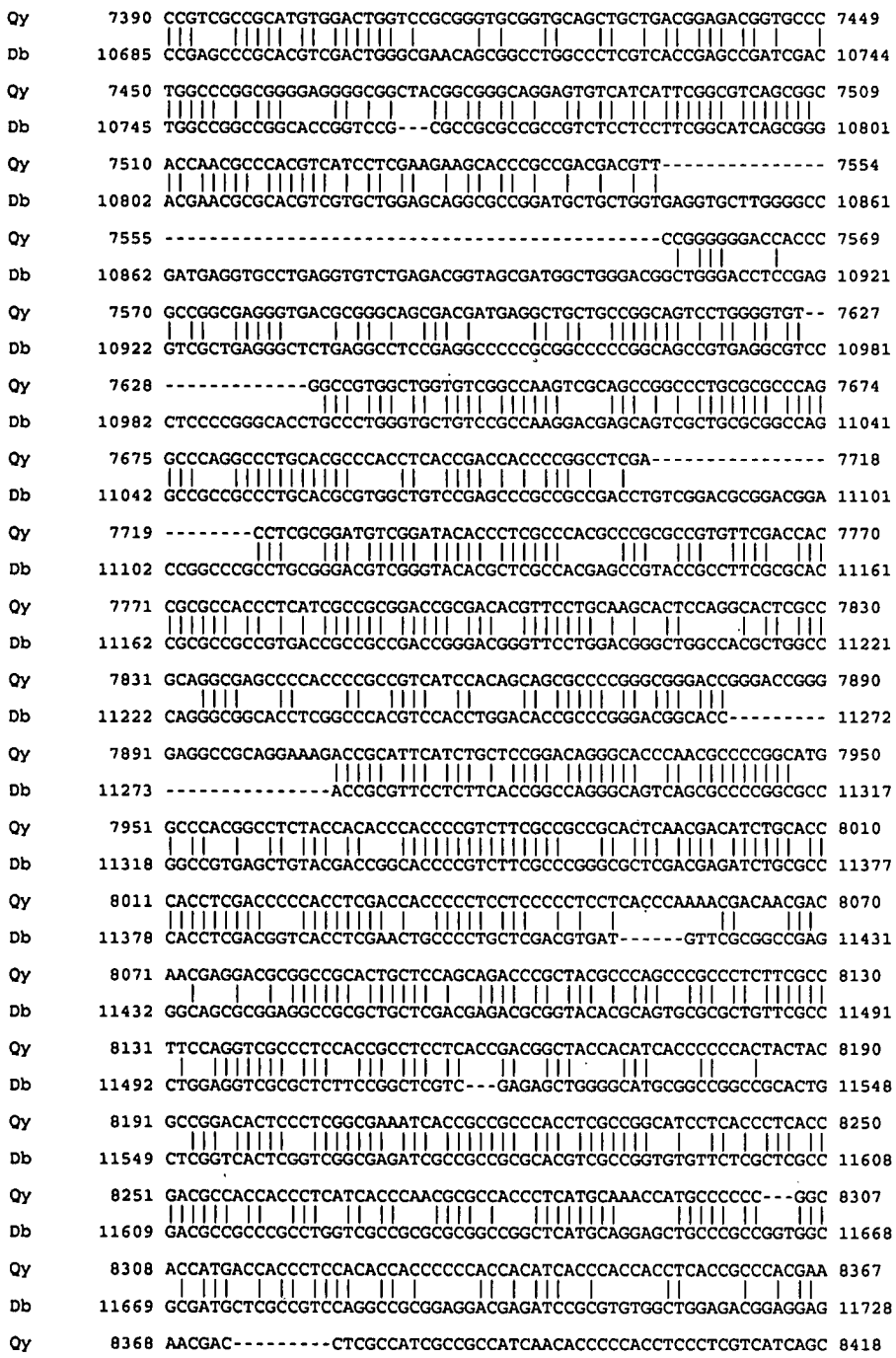
Query Match 9.6%; Score 2951.8; DB 10; Length 36778;
Best Local Similarity 54.4%; Pred. No. 0;
Matches 8436; Conservative 0; Mismatches 5722; Indels 1361; Gaps 69;

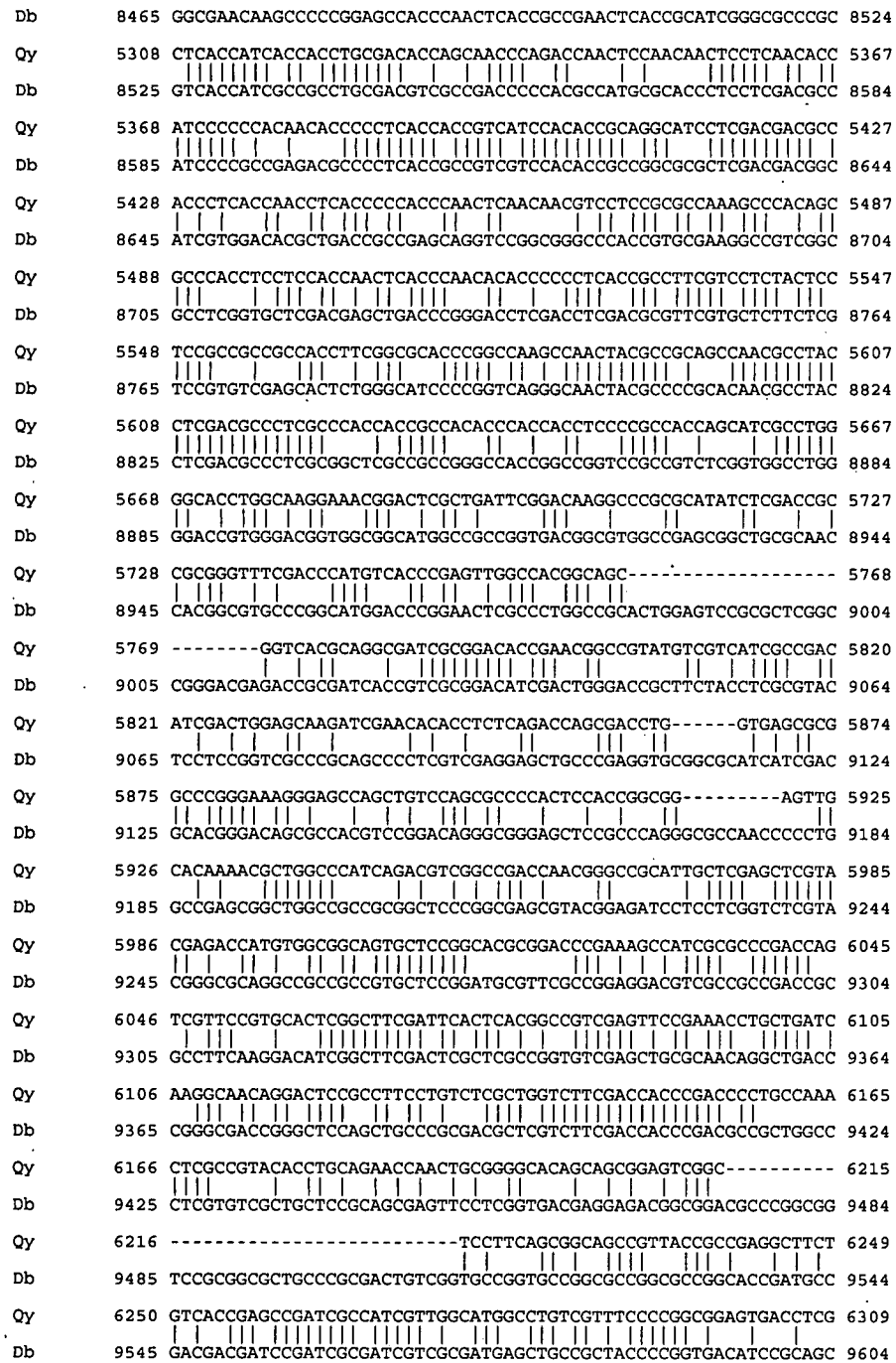


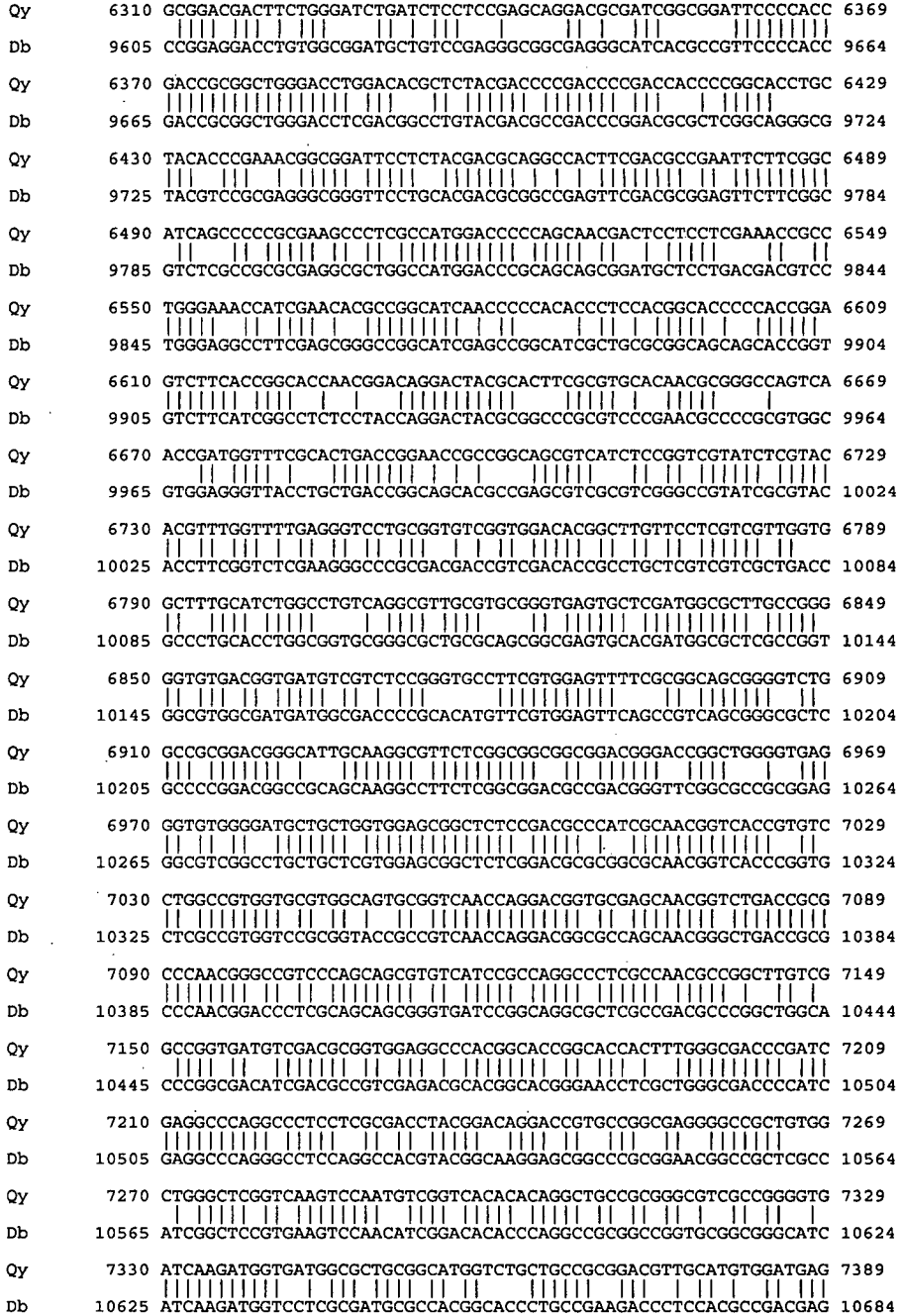

Qy 11453 TCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCC 11512
Db 14970 TCGAGTTCGTCGTCGGCGAGGTCGCCGAAGTACTCGGCCACGCCCGCGGTCACCGGATCG 15029
Qy 11513 CCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCA 11572
Db 15030 ACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGACCGCCGTCGAACTCCGCA 15089
Qy 11573 ACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCA 11632
Db 15090 ACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGCGACCCTGGTCTTCGACCACCCAA 15149
Qy 11633 ACCCCACCACCCTCACCCACCACCTCCACACACAACTCC---------- 11671
Db 15150 GCCCGGCGGCACTCGCCTCCCACCTGGACGCCGAGCTGCCGCGCGGCGCCTCGGACCAGG 15209
Qy 11672 ----------AGCCACAACCGGACAACGCTGTCGC---------- 11696
Db 15210 ACGGAGCCGGGAACCGGAACGGGAACGAGAACGGGACGACGGCGTCCCGGAGCACCGCCG 15269
Qy 11697 -------CCCCGTGTTGGCGGAGCTCGACAAACTCGAATCCGCCCTCTCCGCCCTCGACA 11749
Db 15270 AGACGGACGCGCTGCTGGCACAACTGACCCGCCTGGAAGGCGCCTTGGTGCTGACGGGCC 15329
Qy 11750 AAACCGACAGCGCC---AGCGAAAGAGTCACCCTGCGGCTGAAGTCACTCATGTTGAGG- 11805
Db 15330 TCTCGGACGCCCCCGGGAGCGAAGAAGTCCTGGAGCACCTGCGGTCCCTGCGCTCGATGG 15389
Qy 11806 ----------TGGAACGCACCCCAGCATCCGACAGCCGAAAGCGCTGATG 11845
Db 15390 TCACGGGCGAGACCGGGACCGGGACCGCGTCCGGAGCCCCGGACGGCGCCGGGTCCGGCG 15449
Qy 11846 ACGACGA---------- 11852
Db 15450 CCGAGGACCGGCCCTGGGCGGCCGGGGACGGAGCCGGGGGCGGGAGTGAGGACGGCGCGG 15509
Qy 11853 -------GAAGTTCACATCGGCAACAGAGGCTGAGATTTTCAAATTCATTGACAACGACC 11905
Db 15510 GAGTGCCGGACTTCATGAACGCCTCGGCCGAGGAACTCTTCGGCCTCCTCGACCAGGACC 15569
Qy 11906 TCGGCCTGTCCTGAACCGGACGCCTGCCACTCCGCCCGTATCCG---------- 11949
Db 15570 CCAGCACGGACTGATCCCTGCCGCACGGTCGCCTCCCGCCCCGGACCCCGTCCCGGGCAC 15629
Qy 11950 ----------CTGGGCCCTGCTAGGACGTGAAT 11972
Db 15630 CTCGACTCGAATCACTTCATGCGCGCCTCGGGCGCCTCCAGGAACTCAAGGGGACAGCGT 15689
Qy 11973 GCAATTGGCGAATGAAGCGAAGCTCCTGGAATACCTCAAGCGCGTCACTGCGGACCTGGA 12032
Db 15690 GTCCACGGTGAACGAAGAGAAGTACCTCGACTACCTGCGTCGTGCCACGGCGGACCTCCA 15749
Qy 12033 CCGCACTCGCCGTCGCCTGTACGAGGTGGTCGAGCGTGAGCAGGAGCCGATCGCGATTGT 12092
Db 15750 CGAGGCCCGTGGCCGCCTCCGCGAGCTGGAGGCGAAGGCGGGCGAGCCGGTGGCGATCGT 15809
Qy 12093 GGGGATGGCGTGTCGTTACCCAGGCGGGGCGACGTCACCCACGCGACTGTGGCATCTCGT 12152
Db 15810 CGGCATGGCCTGCCGCCTGCCCGGCGGCGTCGCCTCGCCCGAGGACCTGTGGCGGCTGGT 15869
Qy 12153 CAAGTCCCAGACGGACGCTATCGGGGAGTTCCCGACCGACCGTGGATGGAACCTGGAGCA 12212
Db 15870 GGCCGGCGGCGAGGACGCGATCTCGGAGTTCCCCCAGGACCGCGGCTGGGACGTGGAGGG 15929
Qy 12213 GCTCTACGACCCGGACCCCGACCGCTCAGGAACCAGTTACACGCGCAGCGGAGGGTTTCT 12272
Db 15930 CCTGTACGACCCGAACCCGGAGGCCACGGGCAAGAGTTACGCCCGCGAGGCCGGATTCCT 15989
Qy 12273 CTATGACGCGGGCGACTTCGACGCCGCGTTCTTCGAGTTGTCACCGCGTGAGGCGCTGGC 12332
Db 15990 GTACGAGGCGGGCGAGTTCGACGCCGACTTCTTCGGGATCTCGCCGCGCGAGGCCCTCGC 16049
Qy 12333 AATGGACCCGCAGCAGCGCCTGCTGCTCGAAACCACTTGGGAAACGTTCGAACAGGGCGG 12392
Db 16050 CATGGACCCGCAGCAGCGTCTCCTCCTGGAGGCCTCCTGGGAGGCGTTCGAGCACGCCGG 16109
Qy 12393 AATCGACCCGAGGTCCATGCGCGGAAGCCGGACCGGGGTTTTCGTGGGGATCAATCCGGA 12452
Db 16110 GATCCCGGCGGCCACCGCGCGCGGCACCTCGGTCGGCGTCTTCACCGGCGTGATGTACCA 16169
Qy 12453 GGACTACACCACCGGATACACACATCAGCCCTCAAACGCAGTCGAGGGCTACCTGCTCAC 12512
Db 16170 CGACTACGCCACCCGTCTCACCGAT---GTCCCGGAGGGCATCGAGGGCTACCTGGGCAC 16226
Qy 12513 TGGCAGCGCGGCAAGCATTGCGTCAGGCCGTATCTCCTACAACTTCGGGCTCGAAGGCCC 12572
Db 16227 CGGCAACTCCGGCAGTGTCGCCTCGGGCCGCGTCGCGTACACGCTTGGCCTGGAGGGGCC 16286
Qy 12573 TGCGATCACTATCGACACCGCGTGTTCCTCCTCGCTCGTCGCCCTGCATCTGGCCTGCCA 12632
Db 16287 GGCCGTCACGGTCGACACCGCCTGCTCGTCCTCGCTGGTCGCCCTGCACCTCGCCGTGCA 16346
Qy 12633 AGCGCTCCGGTCCGGTGAATGCACCATGGCGCTCGCAGGCGGCGCCTCCGTCATGGCCAC 12692
Db 16347 GGCCCTGCGCAAGGGCGAGGTCGACATGGCGCTCGCCGGCGGCGTGACGGTCATGTCGAC 16406
Qy 12693 TCCCTTCGTCTTCACCGAGTTCTCTCGCCAGCGGGGCCTGGCCGCAGACGGCCGGTGCAA 12752
Db 16407 GCCCAGCACCTTCGTCGAGTTCAGCCGTCAGCGCGGGCTGGCGCCGGACGGCCGGTCGAA 16466
Qy 12753 GGCGTTTTCGGCGGCGGCGGACGGGACCGGCTGGTCCGAGGGTGTGGGGATGCTGCTGGT 12812
Db 16467 GTCCTTCTCGTCGACGGCCGACGGCACCAGCTGGTCCGAGGGCGTCGGCGTCCTCCTCGT 16526
Qy 12813 GGAGCGGCTCTCCGACGCCCGCCGCAACGGTCACCGTGTCCTGGCCGTCGTCCGCGGCAG 12872
Db 16527 CGAGCGCCTGTCCGACGCGCGTCGCAAGGGCCATCGGATCCTCGCCGTGGTCCGGGGCAC 16586
Qy 12873 CGCCGTCAACCAGGACGGCGCAAGCAACGGCCTGACCGCACCCAACGGTCGTTCACAAGT 12932
Db 16587 CGCCGTCAACCAGGACGGCGCCAGCAGCGGCCTCACGGCTCCGAACGGGCCGTCGCAGCA 16646
Qy 12933 CAAGGTCATCCGCCAGGCTTTGGCCAACGCACACCTCTCCCCTGCCGATGTCGATGCGGT 12992
Db 16647 GCGCGTCATCCGACGTGCCCTGGCGGACGCCCGGCTCACGACCTCCGACGTGGACGTCGT 16706
Qy 12993 GGAGGCCCACGGCACGGGGACCACCCTGGGCGACCCGATCGAGGCTCAAGCCCTCGTCGA 13052
Db 16707 CGAGGCCCACGGCACGGGTACGCGACTCGGCGACCCGATCGAGGCGCAGGCCGTCATCGC 16766
Qy 13053 AGCCTACGGTCAGGACCGCCCCAACGGCCGCCCCCTCTGGCTCGGAACCCTCAAGTCCAA 13112
Db 16767 CACGTACGGGCAGGGCCGTGACGGCGAACAGCCGCTGCGCCTCGGGTCGTTGAAGTCCAA 16826
Qy 13113 CATCGGGCACTCCATGGCCGCTGCGGGTGTGGGCGGGGTCATCAAGATGGTGATGGCGCT 13172
Db 16827 CATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCGTGATCAAGATGGTCCAGGCGAT 16886
Qy 13173 GCGGAATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCGCCGCATGTGGACTG 13232
Db 16887 GCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGAGAAGCCGACGGACCAGGTGGACTG 16946
Qy 13233 GTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCCGGC---GGGGAGGG 13289
Db 16947 GTCCGCGGGCGCGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCGGACAAGGGCGACGG 17006
Qy 13290 GCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAACGCCCACGTCAT 13349
Db 17007 CGGACTGCGCAGGGCCGCGGTCTCCTCCTTCGGCGTCAGCGGGACGAACGCGCACGTCGT 17066
Qy 13350 CCTCGAGGAAGCACCCGCCCACAACATCCCGTCAGACACACCCGCCGACGACGTCCCGGG 13409
Db 17067 GCTCGAAGAGGCCCCGG----------CGGCCGAGG 17092
Qy 13410 AGAATCAGCCGCCGACGAGGATGCCGGTAGTGGCGATGAGGCTGCTGCCGGCAGTCCAGG 13469

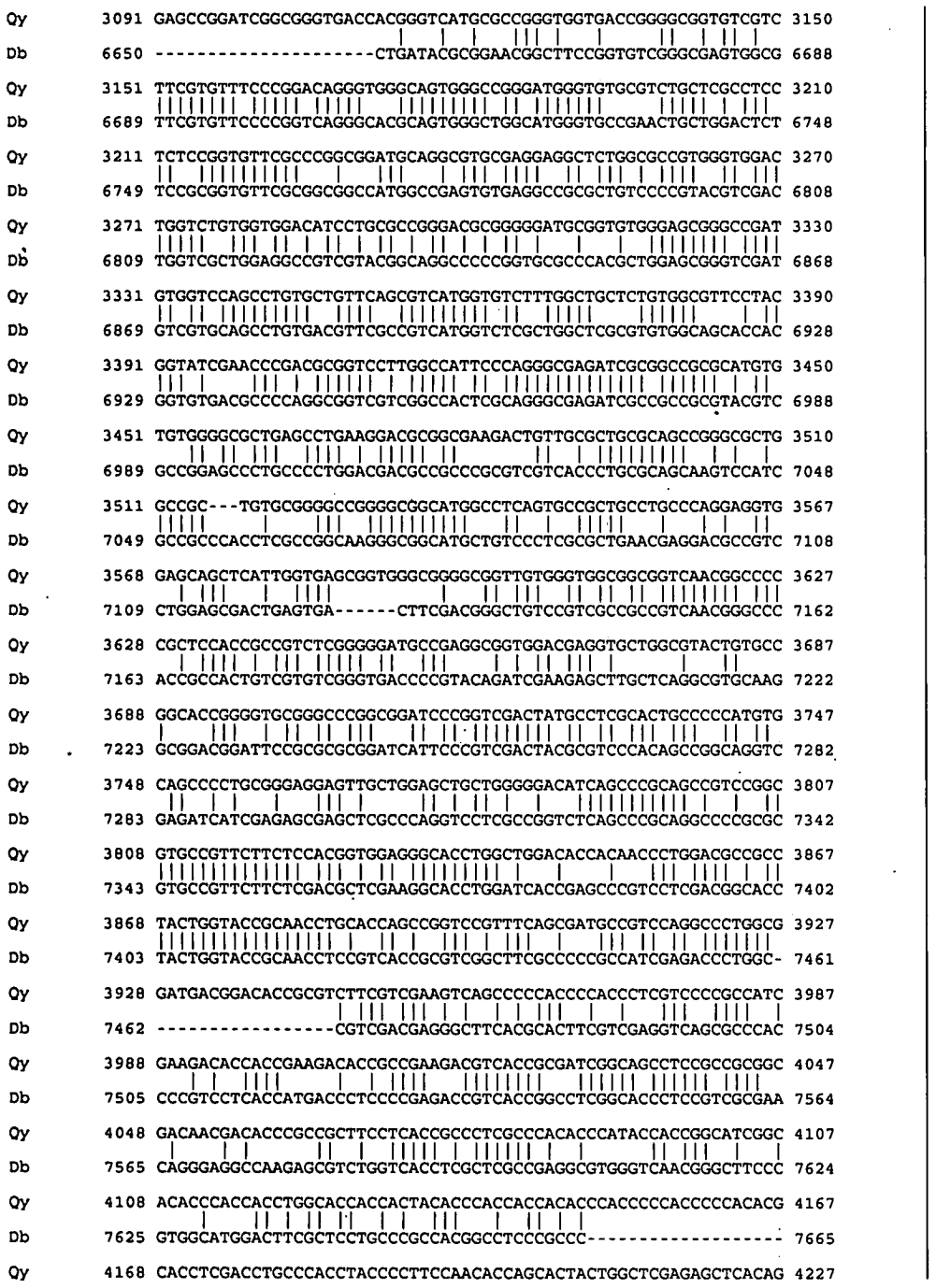


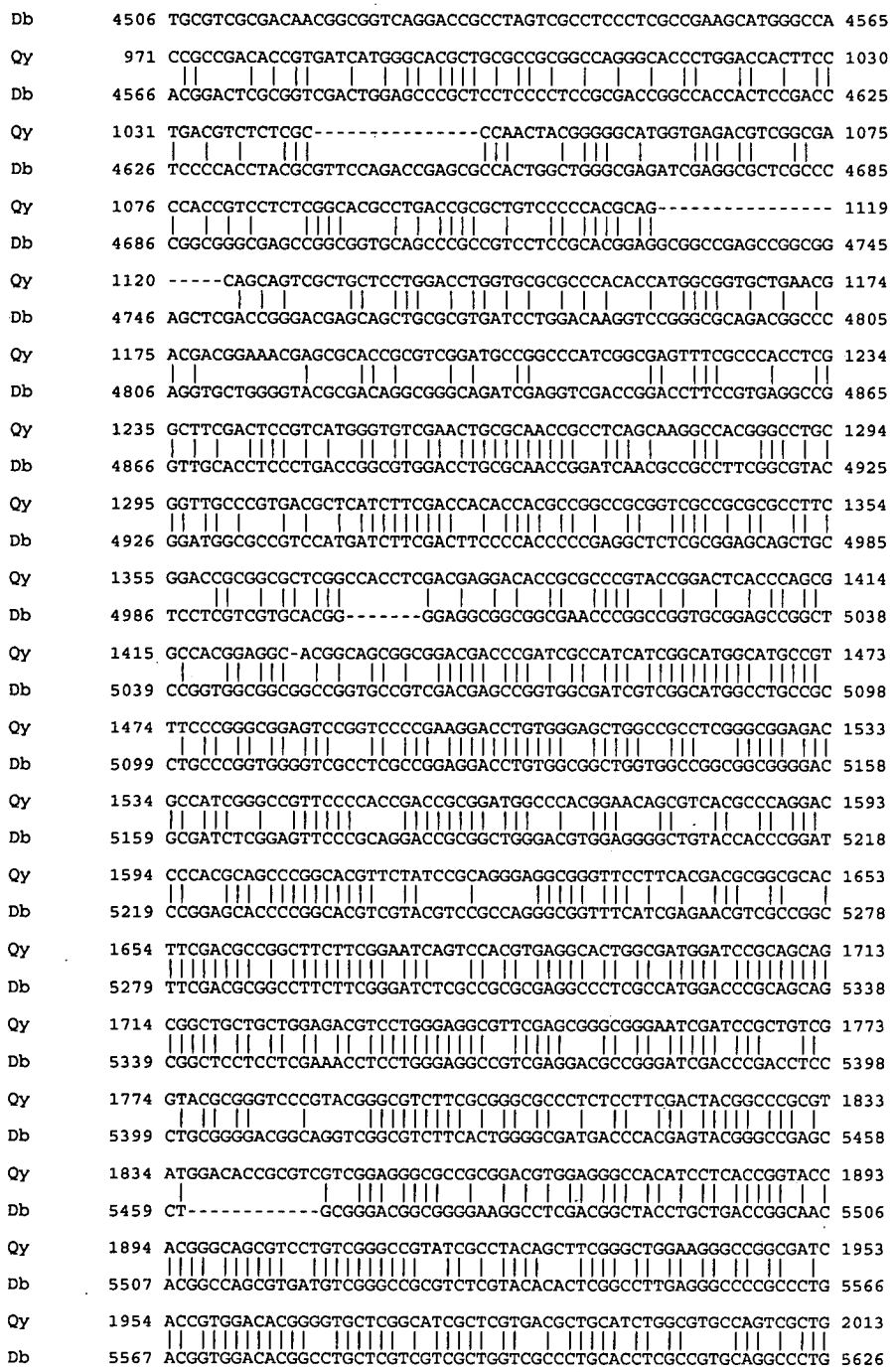


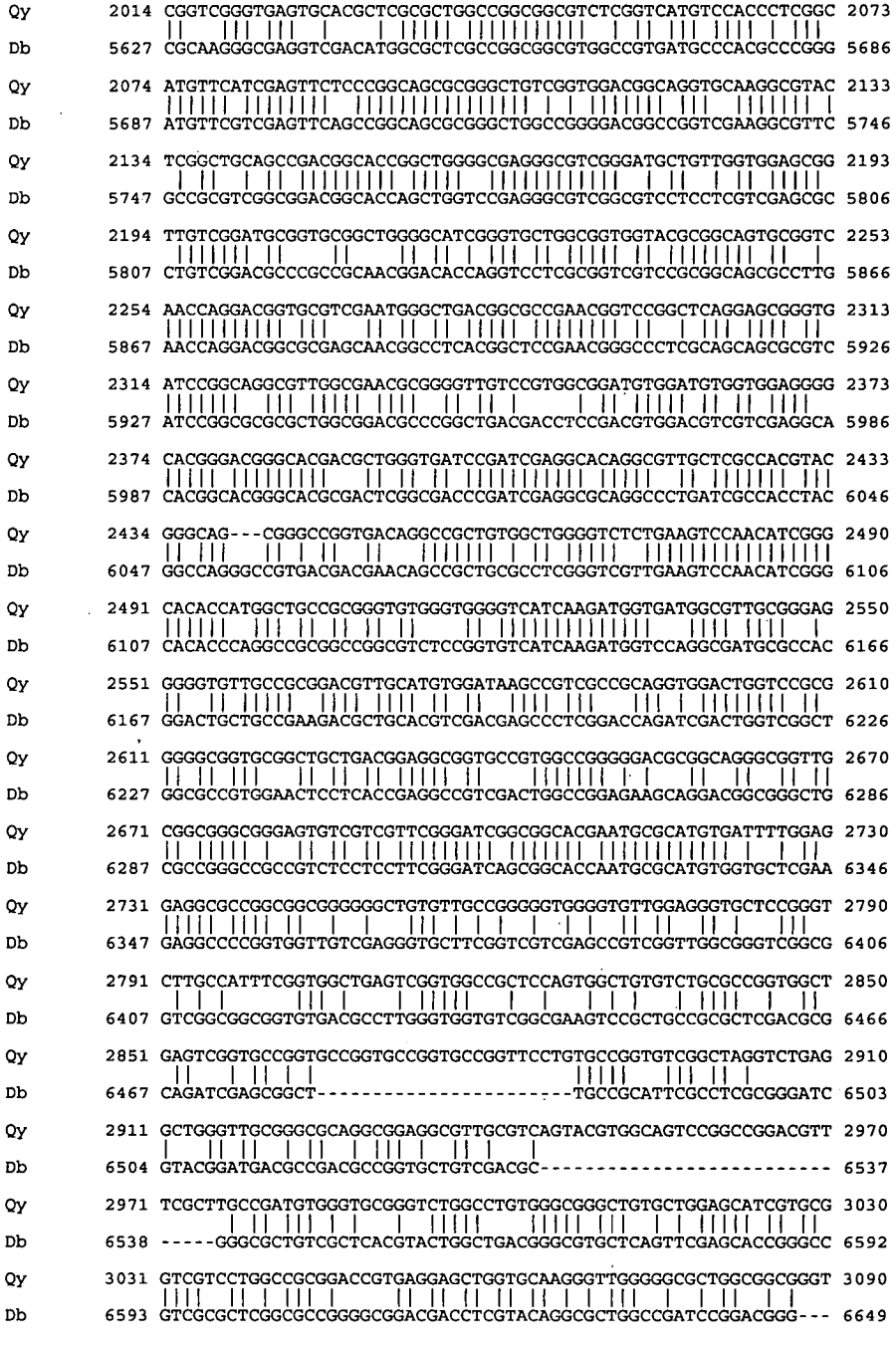

Qy 7390 CCGTCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCC 7449
Db 10685 CCGAGCCCGCACGTCGACTGGGCGAACAGCGGCCTGGCCCTCGTCACCGAGCCGATCGAC 10744
Qy 7450 TGGCCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGC 7509
Db 10745 TGGCCGGCCGGCACCGGTCCG---CGCCGCGCCGCCGTCTCCTCCTTCGGCATCAGCGGG 10801
Qy 7510 ACCAACGCCCACGTCATCCTCGAAGAAGCACCCGCCGACGACGTT---------- 7554
Db 10802 ACGAACGCGCACGTCGTGCTGGAGCAGGCGCCGGATGCTGCTGGTGAGGTGCTTGGGGCC 10861
Qy 7555 ----------CCGGGGGGACCACCC 7569
Db 10862 GATGAGGTGCCTGAGGTGTCTGAGACGGTAGCGATGGCTGGGACGGCTGGGACCTCCGAG 10921
Qy 7570 GCCGGCGAGGGTGACGCGGGCAGCGACGATGAGGCTGCTGCCGGCAGTCCTGGGGTGT-- 7627
Db 10922 GTCGCTGAGGGCTCTGAGGCCTCCGAGGCCCCCGCGGCCCCCGGCAGCCGTGAGGCGTCC 10981
Qy 7628 ----------GGCCGTGGCTGGTGTCGGCCAAGTCGCAGCCGGCCCTGCGCGCCCAG 7674
Db 10982 CTCCCCGGGCACCTGCCCTGGGTGCTGTCCGCCAAGGACGAGCAGTCGCTGCGCGGCCAG 11041
Qy 7675 GCCCAGGCCCTGCACGCCCACCTCACCGACCACCCCGGCCTCGA---------- 7718
Db 11042 GCCGCCGCCCTGCACGCGTGGCTGTCCGAGCCCGCCGCCGACCTGTCGGACGCGGACGGA 11101
Qy 7719 --------CCTCGCGGATGTCGGATACACCCTCGCCCACGCCCGCGCCGTGTTCGACCAC 7770
Db 11102 CCGGCCCGCCTGCGGGACGTCGGGTACACGCTCGCCACGAGCCGTACCGCCTTCGCGCAC 11161
Qy 7771 CGCGCCACCCTCATCGCCGCGGACCGCGACACGTTCCTGCAAGCACTCCAGGCACTCGCC 7830
Db 11162 CGCGCCGCCGTGACCGCCGCCGACCGGGACGGGTTCCTGGACGGGCTGGCCACGCTGGCC 11221
Qy 7831 GCAGGCGAGCCCCACCCCGCCGTCATCCACAGCAGCGCCCCGGGCGGGACCGGGACCGGG 7890
Db 11222 CAGGGCGGCACCTCGGCCCACGTCCACCTGGACACCGCCCGGGACGGCACC--------- 11272
Qy 7891 GAGGCCGCAGGAAAGACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATG 7950
Db 11273 ----------ACCGCGTTCCTCTTCACCGGCCAGGGCAGTCAGCGCCCCGGCGCC 11317
Qy 7951 GCCCACGGCCTCTACCACACCCACCCCGTCTTCGCCGCCGCACTCAACGACATCTGCACC 8010
Db 11318 GGCCGTGAGCTGTACGACCGGCACCCCGTCTTCGCCCGGGCGCTCGACGAGATCTGCGCC 11377
Qy 8011 CACCTCGACCCCCACCTCGACCACCCCCTCCTCCCCCTCCTCACCCAAAACGACAACGAC 8070
Db 11378 CACCTCGACGGTCACCTCGAACTGCCCCTGCTCGACGTGAT------GTTCGCGGCCGAG 11431
Qy 8071 AACGAGGACGCGGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCC 8130
Db 11432 GGCAGCGCGGAGGCCGCGCTGCTCGACGAGACGCGGTACACGCAGTGCGCGCTGTTCGCC 11491
Qy 8131 TTCCAGGTCGCCCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTAC 8190
Db 11492 CTGGAGGTCGCGCTCTTCCGGCTCGTC---GAGAGCTGGGGCATGCGGCCGGCCGCACTG 11548
Qy 8191 GCCGGACACTCCCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACC 8250
Db 11549 CTCGGTCACTCGGTCGGCGAGATCGCCGCCGCGCACGTCGCCGGTGTGTTCTCGCTCGCC 11608
Qy 8251 GACGCCACCACCCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCC---GGC 8307
Db 11609 GACGCCGCCCGCCTGGTCGCCGCGCGCGGCCGGCTCATGCAGGAGCTGCCCGCCGGTGGC 11668
Qy 8308 ACCATGACCACCCTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAA 8367
Db 11669 GCGATGCTCGCCGTCCAGGCCGCGGAGGACGAGATCCGCGTGTGGCTGGAGACGGAGGAG 11728
Qy 8368 AACGAC---------CTCGCCATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGC 8418
Db 11729 CGGTACGCGGGACGTCTGGACGTCGCCGCCGTCAACGGCCCCGAGGCCGCCGTCCTGTCC 11788
Qy 8419 GGCACCCCCCACACCGTCCAACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACC 8478
Db 11789 GGCGACGCGGACGCGGCGCGGGAGGCGGAGGCGTACTGGTCCGGGCTCGGCCGCAGGACC 11848
Qy 8479 AAAACCCTCCCCACCAACCACGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAA 8538
Db 11849 CGCGCGCTGCGGGTCAGCCACGCCTTCCACTCCGCGCACATGGACGGCATGCTCGACGGG 11908
Qy 8539 CTCCACCAGCACACCCAAACCCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCC 8598
Db 11909 TTCCGCGCCGTCCTGGAGACGGTGGAGTTCCGGCGCCCCTCCCTGACCGTGGTCTCGAAC 11968
Qy 8599 AACA----------CCCCACCCGACCAACTCCTCACCCCCCACTACTGGACCCAACAA 8646
Db 11969 GTCACCGGCCTGGCCGCCGGCCCGGACGACCTGTGCGACCCCGAGTACTGGGTCCGGCAC 12028
Qy 8647 GCCCGCAACACCGTCGACTACGCCACCACCACCCAAACCCTCCACCAACACGGCGTCACC 8706
Db 12029 GTCCGCGGCACCGTCCGCTTCCTCGACGGCGTCCGTGTCCTGCGCGACCTCGGCGTGCGG 12088
Qy 8707 ACCTACATCGAACTCGGACCCGACAACACCCTCACCACCCTCACCCACCACAACCTCCCC 8766
Db 12089 ACCTGCCTGGAGCTGGGCCCCGACGGGGTCCTCACCGCCATGGCGGCCGACGGCCTCGCG 12148
Qy 8767 AACCCCCCCACCACCACCCTCACCCTCACCCAC---------- 8799
Db 12149 GACACCCCCGCGGATTCCGCTGCCGGCTCCCCCGTCGGCTCTCCCGCCGGCTCTCCCGCC 12208
Qy 8800 ----------CCCCACCACCACCCCCAAACCCACCTCCTCACCAAC 8835
Db 12209 GACTCCGCCGCCGGCGCGCTCCGGCCCCGGCCGCTGCTCGTGGCGCTGCTGCGCCGCAAG 12268
Qy 8836 CTCGCCAAAACCACCACCACCTGGCACCCCCACCACTACACCCA---CCACGACAACCAA 8892
Db 12269 CGGTCGGAGACCGAGACCGTCGCGGACGCCCTCGGCAGGGCGCACGCCCACGGCACCGGA 12328
Qy 8893 CCCCACACCCACAC----------CCACCTCGACCTCCCCACCTAC 8928
Db 12329 CCCGACTGGCACGCCTGGTTCGCCGGCTCCGGGGCGCACCGCGTGGACCTGCCCACGTAC 12388
Qy 8929 CCCTTCCAACACCACCACTACTGGCTCGAAAGCACACAGCCCGGTGCCGGCAACGTGTCA 8988
Db 12389 TCCTTCCGGCGCGACCGCTACTGGCTGGACGCCCCGGCGGCCGACACCGCGGTGGACACC 12448
Qy 8989 GCAG---CCGGACTCGACCCCACCGAACACCCCCTACTCGGCGCCACATTGGAACTGGCG 9045
Db 12449 GCCGGCCTCGGTCTCGGCACCGCCGACCACCCGCTGCTCGGCGCCGTGGTCAGCCTTCCG 12508
Qy 9046 ACTGACGGTGGAGCGCTTCTTGCAGGGCGCTTGTCTTTGAGGTCGCATCCGTGGCTGGCT 9105
Db 12509 GACCGGGACGGCCTGCTGCTCACCGGCCGCCTCTCCCTGCGCACCCACCCGTGGCTCGCG 12568
Qy 9106 GACCATGCCGTCGGCGGCACGGTGCTGCTGTCGGGCGCCACCTTCCTCGAACTCGCCCTT 9165
Db 12569 GACCACGCCGTCCTGGGGAGCGTCCTGCTCCCCGGCGCCGCGATGGTCGAACTCGCCGCG 12628
Qy 9166 CATGCGGGCACATACGTGGGCTGCGACCGAGTGGATGAGCTGACGCTGCATGCGCCGCTG 9225
Db 12629 CACGCTGCGGAGTCCGCCGGTCTGCGTGACGTGCGGGAGCTGACCCTCCTTGAACCGCTG 12688
Qy 9226 GTGGTTCCTGTGGATGGGGGTGTGAGTGTGCAGGTTGGGGTTGCGGCTGCGGATGG---- 9281
Db 12689 GTACTGCCCGAGCACGGTGGCGTCGAGCTGCGCGTGACGGTCGGGGCGCCGGCCGGAGAG 12748
Qy 9282 ---GGAGGGGCGGCGTTTGGTGAGTGTGTATGCGCGGGGTGGGAGTGCTTGTGG--TGGG 9336
Db 12749 CCCGGTGGCGAGTCGGCCGGGGACGGCGCACGGCCCGTCTCCCTCCACTCGCGGCTCGCC 12808
Qy 9337 GGTGGTGCGTCGGGTGGGGTGTGGACGTGTCATGCCTCGGGGGTGCTGGTTGAGGCTGCT 9396






Qy 3091 GAGCCGGATCGGCGGGTGACCACGGGTCATGCGCCGGGTGGTGACCGGGGCGGTGTCGTC 3150
Db 6650 ----------CTGATACGCGGAACGGCTTCCGGTGTCGGGCGAGTGGCG 6688
Qy 3151 TTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCC 3210
Db 6689 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCTGGCATGGGTGCCGAACTGCTGGACTCT 6748
Qy 3211 TCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGAC 3270
Db 6749 TCCGCGGTGTTCGCGGCGGCCATGGCCGAGTGTGAGGCCGCGCTGTCCCCGTACGTCGAC 6808
Qy 3271 TGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGAT 3330
Db 6809 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 6868
Qy 3331 GTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTAC 3390
Db 6869 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 6928
Qy 3391 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 3450
Db 6929 GGTGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 6988
Qy 3451 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 3510
Db 6989 GCCGGAGCCCTGCCCCTGGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 7048
Qy 3511 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 3567
Db 7049 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAACGAGGACGCCGTC 7108
Qy 3568 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 3627
Db 7109 CTGGAGCGACTGAGTGA------CTTCGACGGGCTGTCCGTCGCCGCCGTCAACGGGCCC 7162
Qy 3628 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 3687
Db 7163 ACCGCCACTGTCGTGTCGGGTGACCCCGTACAGATCGAAGAGCTTGCTCAGGCGTGCAAG 7222
Qy 3688 GGCACCGGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 3747
Db 7223 GCGGACGGATTCCGCGCGCGGATCATTCCCGTCGACTACGCGTCCCACAGCCGGCAGGTC 7282
Qy 3748 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 3807
Db 7283 GAGATCATCGAGAGCGAGCTCGCCCAGGTCCTCGCCGGTCTCAGCCCGCAGGCCCCGCGC 7342
Qy 3808 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 3867
Db 7343 GTGCCGTTCTTCTCGACGCTCGAAGGCACCTGGATCACCGAGCCCGTCCTCGACGGCACC 7402
Qy 3868 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 3927
Db 7403 TACTGGTACCGCAACCTCCGTCACCGCGTCGGCTTCGCCCCCGCCATCGAGACCCTGGC- 7461
Qy 3928 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 3987
Db 7462 ----------CGTCGACGAGGGCTTCACGCACTTCGTCGAGGTCAGCGCCCAC 7504
Qy 3988 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 4047
Db 7505 CCCGTCCTCACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGCACCCTCCGTCGCGAA 7564
Qy 4048 GACAACGACACCCGCCGCTTCCTCACCGCCCTCGCCCACACCCATACCACCGGCATCGGC 4107
Db 7565 CAGGGAGGCCAAGAGCGTCTGGTCACCTCGCTCGCCGAGGCGTGGGTCAACGGGCTTCCC 7624
Qy 4108 ACACCCACCACCTGGCACCACCACTACACCCACCACCACACCCACCCCCACCCCCACACG 4167
Db 7625 GTGGCATGGACTTCGCTCCTGCCCGCCACGGCCTCCCGCCC---------- 7665
Qy 4168 CACCTCGACCTGCCCACCTACCCCTTCCAACACCAGCACTACTGGCTCGAGAGCTCACAG 4227
Db 7666 -----CGGTCTGCCCACCTACGCCTTCCAGGCCGAGCGCTACTGGCTCGAGAACACTCCC 7720
Qy 4228 CCGGGTGCCGGATCCGGTTCGGGTGCCGGTGCCGGTTCGGGTGCCGGTTCCGGGCGGGCA 4287
Db 7721 GCCGC----------CCTGGCCACCG 7736
Qy 4288 GGGACTGCGGGCGGGACGGCAGAGGTGGAGTCGCGGTTCTGGGACGCGGTGGCCCGCCAG 4347
Db 7737 GCGACGACTGGCGCTACCGCATCGACTG-GAAGCGCCTCCCGGCCGCCGAGGGGTCCGAG 7795
Qy 4348 GACCTGGAAACGGTCGCGACCACACTCGCCGTGCCCCCCTCCGCCGGCCTGGACACGGTG 4407
Db 7796 CGCACCGGCCTGTCCGGCCGCTGGCTCGCCGTCACGCCGGAGGACCACTCCGCGCAGGCC 7855
Qy 4408 GTGCCCGCACTCTCCGCCTGGCACCGCCACCAACACGACCAAGCCCGCATCAACACCTGG 4467
Db 7856 GCCGCCGTGCTCACCGCGCTGGTCGACGCCGGGGCGAAGGTCGAGGTGCTGACGGCCGGG 7915
Qy 4468 ACCTACCAGGAAACCTGGAAACCCCTCACCCTCCCCACCACCCACCAACCCCACCAAACC 4527
Db 7916 GCGGACGACGACCGTGAGGCCCTCGCCGCCCGGCTCACCGCACTGACG---------- 7963
Qy 4528 TGGCTCATCGCCATCCCCGAAACCCAGACCCACCACCCCCACATCACCAACATCCTCACC 4587
Db 7964 ---------- 7963
Qy 4588 AACCTCCACCACCACGGCATCACCCCCATCCCCCTCACCCTCAACCACACCCACACCAAC 4647
Db 7964 ---------- 7963
Qy 4648 CCCCAACACCTCCACCACACCCTCCACCACACCCGACAACAAGCCCAAAACCACACCACC 4707
Db 7964 ----------ACCGGT 7969
Qy 4708 GGAGCCATCACCGGCCTGCTCTCCCTCCTCGCCCTCGACGAAACACCCCACCCCCACCAC 4767
Db 7970 GACGGCTTCACCGGCGTGGTCTCGCTCCTCG---------- 8000
Qy 4768 CCCCACACACCCACCGGCACCCTCCTCAACCTCACCCTCACCCAAACCCACACCCAAACC 4827
Db 8001 ----------ACGGACTCGTACCGCAGGTCGCCTGGGTCCAGGCGCTCGGCGACGCC 8047
Qy 4828 CACCCACCAACCCCCCTCTGGTACGCCACCACCAACGCCACCACCACCCACCCCAACGAC 4887
Db 8048 GGAATCAAGGCGCCCCTGTGGTCCGTCACCCAGGGCGCGGTCTCCGTCGGACGTCTCGAC 8107
Qy 4888 CCCCTCACACACCCCACCCAAGCCCAAACCTGGGGACTCGCCCGCACCACCCTCCTCGAA 4947
Db 8108 ACCCCCGCCGACCCCGACCGGGCCATGCTCTGGGGCCTCGGCCGCGTCGTCGCCCTTGAG 8167
Qy 4948 CACCCCACCCACACCGCCGGAATCATCGACCTCCCCACCACCCCCACCCCCCACACCCTC 5007
Db 8168 CACCCCGAACGCTGGGCCGGCCTCGTCGACCTCCCCGCCCAGCCCGATGCCGCCGCCCTC 8227
Qy 5008 CAGCACCTCACCCAAACCCTCACCCAACCCCACCACCAAACCCAACTCGCCATCCGCACC 5067
Db 8228 GCCCACCTCGTCACCGCACTCTCCGGCGCCACCGGCGAGGACCAGATCGCCATCCGCACC 8287
Qy 5068 ACCGGCACCCACACCCGCCGCCTCACCCCCACCACCCTCACCCCCACACACCAACCACCC 5127
Db 8288 ACCGGACTCCACGCCCGCCGCCTCGCCCGCGCACCCCTCCACGGACGTCGGCCCACCCGC 8347
Qy 5128 ACCCCCACCCCCCACGGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCAC 5187
Db 8348 GACTGGCAGCCCCACGGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCA- 8406
Qy 5188 CTCACCCACCACCTCACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACC 5247
Db 8407 --CGCCGCACGCTGGATGGCCCACCACGGAGCCGAACACCTCCTCCTCGTCAGCCGCAGC 8464
Qy 5248 GGCCCCCACACCCCCCACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCAC 5307


Db 4506 TGCGTCGCGACAACGGCGGTCAGGACCGCCTAGTCGCCTCCCTCGCCGAAGCATGGGCCA 4565
Qy 971 CCGCCGACACCGTGATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCC 1030
Db 4566 ACGGACTCGCGGTCGACTGGAGCCCGCTCCTCCCCTCCGCGACCGGCCACCACTCCGACC 4625
Qy 1031 TGACGTCTCTCGC----------CCAACTACGGGGGCATGGTGAGACGTCGGCGA 1075
Db 4626 TCCCCACCTACGCGTTCCAGACCGAGCGCCACTGGCTGGGCGAGATCGAGGCGCTCGCCC 4685
Qy 1076 CCACCGTCCTCTCGGCACGCCTGACCGCGCTGTCCCCCACGCAG---------- 1119
Db 4686 CGGCGGGCGAGCCGGCGGTGCAGCCCGCCGTCCTCCGCACGGAGGCGGCCGAGCCGGCGG 4745
Qy 1120 -----CAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGCGGTGCTGAACG 1174
Db 4746 AGCTCGACCGGGACGAGCAGCTGCGCGTGATCCTGGACAAGGTCCGGGCGCAGACGGCCC 4805
Qy 1175 ACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTTCGCCCACCTCG 1234
Db 4806 AGGTGCTGGGGTACGCGACAGGCGGGCAGATCGAGGTCGACCGGACCTTCCGTGAGGCCG 4865
Qy 1235 GCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGC 1294
Db 4866 GTTGCACCTCCCTGACCGGCGTGGACCTGCGCAACCGGATCAACGCCGCCTTCGGCGTAC 4925
Qy 1295 GGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGCCGCGCGCCTTC 1354
Db 4926 GGATGGCGCCGTCCATGATCTTCGACTTCCCCACCCCCGAGGCTCTCGCGGAGCAGCTGC 4985
Qy 1355 GGACCGCGGCGCTCGGCCACCTCGACGAGGACACCGCGCCCGTACCGGACTCACCCAGCG 1414
Db 4986 TCCTCGTCGTGCACGG-------GGAGGCGGCGGCGAACCCGGCCGGTGCGGAGCCGGCT 5038
Qy 1415 GCCACGGAGGC-ACGGCAGCGGCGGACGACCCGATCGCCATCATCGGCATGGCATGCCGT 1473
Db 5039 CCGGTGGCGGCGGCCGGTGCCGTCGACGAGCCGGTGGCGATCGTCGGCATGGCCTGCCGC 5098
Qy 1474 TTCCCGGGCGGAGTCCGGTCCCCGAAGGACCTGTGGGAGCTGGCCGCCTCGGGCGGAGAC 1533
Db 5099 CTGCCCGGTGGGGTCGCCTCGCCGGAGGACCTGTGGCGGCTGGTGGCCGGCGGCGGGGAC 5158
Qy 1534 GCCATCGGGCCGTTCCCCACCGACCGCGGATGGCCCACGGAACAGCGTCACGCCCAGGAC 1593
Db 5159 GCGATCTCGGAGTTCCCGCAGGACCGCGGCTGGGACGTGGAGGGGCTGTACCACCCGGAT 5218
Qy 1594 CCCACGCAGCCCGGCACGTTCTATCCGCAGGGAGGCGGGTTCCTTCACGACGCGGCGCAC 1653
Db 5219 CCGGAGCACCCCGGCACGTCGTACGTCCGCCAGGGCGGTTTCATCGAGAACGTCGCCGGC 5278
Qy 1654 TTCGACGCCGGCTTCTTCGGAATCAGTCCACGTGAGGCACTGGCGATGGATCCGCAGCAG 1713
Db 5279 TTCGACGCGGCCTTCTTCGGGATCTCGCCGCGCGAGGCCCTCGCCATGGACCCGCAGCAG 5338
Qy 1714 CGGCTGCTGCTGGAGACGTCCTGGGAGGCGTTCGAGCGGGCGGGAATCGATCCGCTGTCG 1773
Db 5339 CGGCTCCTCCTCGAAACCTCCTGGGAGGCCGTCGAGGACGCCGGGATCGACCCGACCTCC 5398
Qy 1774 GTACGCGGGTCCCGTACGGGCGTCTTCGCGGGCGCCCTCTCCTTCGACTACGGCCCGCGT 1833
Db 5399 CTGCGGGGACGGCAGGTCGGCGTCTTCACTGGGGCGATGACCCACGAGTACGGGCCGAGC 5458
Qy 1834 ATGGACACCGCGTCGTCGGAGGGCGCCGCGGACGTGGAGGGCCACATCCTCACCGGTACC 1893
Db 5459 CT------------GCGGGACGGCGGGGAAGGCCTCGACGGCTACCTGCTGACCGGCAAC 5506
Qy 1894 ACGGGCAGCGTCCTGTCGGGCCGTATCGCCTACAGCTTCGGGCTGGAAGGGCCGGCGATC 1953
Db 5507 ACGGCCAGCGTGATGTCGGGCCGCGTCTCGTACACACTCGGCCTTGAGGGCCCCGCCCTG 5566
Qy 1954 ACCGTGGACACGGGGTGCTCGGCATCGCTCGTGACGCTGCATCTGGCGTGCCAGTCGCTG 2013
Db 5567 ACGGTGGACACGGCCTGCTCGTCGTCGCTGGTCGCCCTGCACCTCGCCGTGCAGGCCCTG 5626


Qy 2014 CGGTCGGGTGAGTGCACGCTCGCGCTGGCCGGCGGCGTCTCGGTCATGTCCACCCTCGGC 2073
Db 5627 CGCAAGGGCGAGGTCGACATGGCGCTCGCCGGCGGCGTGGCCGTGATGCCCACGCCCGGG 5686
Qy 2074 ATGTTCATCGAGTTCTCCCGGCAGCGCGGGCTGTCGGTGGACGGCAGGTGCAAGGCGTAC 2133
Db 5687 ATGTTCGTCGAGTTCAGCCGGCAGCGCGGGCTGGCCGGGGACGGCCGGTCGAAGGCGTTC 5746
Qy 2134 TCGGCTGCAGCCGACGGCACCGGCTGGGGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGG 2193
Db 5747 GCCGCGTCGGCGGACGGCACCAGCTGGTCCGAGGGCGTCGGCGTCCTCCTCGTCGAGCGC 5806
Qy 2194 TTGTCGGATGCGGTGCGGCTGGGGCATCGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTC 2253
Db 5807 CTGTCGGACGCCCGCCGCAACGGACACCAGGTCCTCGCGGTCGTCCGCGGCAGCGCCTTG 5866
Qy 2254 AACCAGGACGGTGCGTCGAATGGGCTGACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTG 2313
Db 5867 AACCAGGACGGCGCGAGCAACGGCCTCACGGCTCCGAACGGGCCCTCGCAGCAGCGCGTC 5926
Qy 2314 ATCCGGCAGGCGTTGGCGAACGCGGGGTTGTCCGTGGCGGATGTGGATGTGGTGGAGGGG 2373
Db 5927 ATCCGGCGCGCGCTGGCGGACGCCCGGCTGACGACCTCCGACGTGGACGTCGTCGAGGCA 5986
Qy 2374 CACGGGACGGGCACGACGCTGGGTGATCCGATCGAGGCACAGGCGTTGCTCGCCACGTAC 2433
Db 5987 CACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCGCAGGCCCTGATCGCCACCTAC 6046
Qy 2434 GGGCAG---CGGGCCGGTGACAGGCCGCTGTGGCTGGGGTCTCTGAAGTCCAACATCGGG 2490
Db 6047 GGCCAGGGCCGTGACGACGAACAGCCGCTGCGCCTCGGGTCGTTGAAGTCCAACATCGGG 6106
Qy 2491 CACACCATGGCTGCCGCGGGTGTGGGTGGGGTCATCAAGATGGTGATGGCGTTGCGGGAG 2550
Db 6107 CACACCCAGGCCGCGGCCGGCGTCTCCGGTGTCATCAAGATGGTCCAGGCGATGCGCCAC 6166
Qy 2551 GGGGTGTTGCCGCGGACGTTGCATGTGGATAAGCCGTCGCCGCAGGTGGACTGGTCCGCG 2610
Db 6167 GGACTGCTGCCGAAGACGCTGCACGTCGACGAGCCCTCGGACCAGATCGACTGGTCGGCT 6226
Qy 2611 GGGGCGGTGCGGCTGCTGACGGAGGCGGTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTG 2670
Db 6227 GGCGCCGTGGAACTCCTCACCGAGGCCGTCGACTGGCCGGAGAAGCAGGACGGCGGGCTG 6286
Qy 2671 CGGCGGGCGGGAGTGTCGTCGTTCGGGATCGGCGGCACGAATGCGCATGTGATTTTGGAG 2730
Db 6287 CGCCGGGCCGCCGTCTCCTCCTTCGGGATCAGCGGCACCAATGCGCATGTGGTGCTCGAA 6346
Qy 2731 GAGGCGCCGGCGGCGGGGGGCTGTGTTGCCGGGGGTGGGGTGTTGGAGGGTGCTCCGGGT 2790
Db 6347 GAGGCCCCGGTGGTTGTCGAGGGTGCTTCGGTCGTCGAGCCGTCGGTTGGCGGGTCGGCG 6406
Qy 2791 CTTGCCATTTCGGTGGCTGAGTCGGTGGCCGCTCCAGTGGCTGTGTCTGCGCCGGTGGCT 2850
Db 6407 GTCGGCGGCGGTGTGACGCCTTGGGTGGTGTCGGCGAAGTCCGCTGCCGCGCTCGACGCG 6466
Qy 2851 GAGTCGGTGCCGGTGCCGGTGCCGGTGCCGGTTCCTGTGCCGGTGTCGGCTAGGTCTGAG 2910
Db 6467 CAGATCGAGCGGCT----------TGCCGCATTCGCCTCGCGGGATC 6503
Qy 2911 GCTGGGTTGCGGGCGCAGGCGGAGGCGTTGCGTCAGTACGTGGCAGTCCGGCCGGACGTT 2970
Db 6504 GTACGGATGACGCCGACGCCGGTGCTGTCGACGC---------- 6537
Qy 2971 TCGCTTGCCGATGTGGGTGCGGGTCTGGCCTGTGGGCGGGCTGTGCTGGAGCATCGTGCG 3030
Db 6538 -----GGGCGCTGTCGCTCACGTACTGGCTGACGGGCGTGCTCAGTTCGAGCACCGGGCC 6592
Qy 3031 GTCGTCCTGGCCGCGGACCGTGAGGAGCTGGTGCAAGGGTTGGGGGCGCTGGCGGCGGGT 3090
Db 6593 GTCGCGCTCGGCGCCGGGGCGGACGACCTCGTACAGGCGCTGGCCGATCCGGACGGG--- 6649

```
Qy      14064 CCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCAC 14123
Db      17673 GCAGGGCGAGATCGCCGCCGCGTACGTCGCCGGTGCCCTCACCCTCGACGACGCCGCCCG 17732

Qy      14124 CCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCAC---CATGACCAC 14180
Db      17733 CGTCGTCACCCTGCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAAGGGCGGCATGAT 17792

Qy      14181 CCTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGC 14240
Db      17793 CTCCCTCGCCCTCAGCGAGGAAGCCACCCGGCAGCGCATCGAGAACCTCCACGGACTGTC 17852

Qy      14241 CATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCA 14300
Db      17853 GATCGCCGCCGTCAACGGCCCCACCGCCACCGTGGTTTCGGGCGACCCCACCCAGATCCA 17912

Qy      14301 ACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCA 14360
Db      17913 AGAGCTCGCTCAGGCGTGTGAGGCCGACGGGGTCCGCGCACGGATCATCCCCGTCGACTA 17972

Qy      14361 CGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAAC 14420
Db      17973 CGCCTCCCACAGCGCCCACGTCGAGACCATCGAGAGCGAACTCGCCGAGGTCCTCGCCGG 18032

Qy      14421 CCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACT 14480
Db      18033 GCTCAGCCCGCGGACACCTGAGGTGCCGTTCTTCTCGACACTCGAAGGCGCCTGGATCAC 18092

Qy      14481 CCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCAC 14540
Db      18093 CGAGCCGGTGCTCGACGGCACCTACTGGTACCGCAACCTCCGCCACCGCGTCGGCTTCGC 18152

Qy      14541 CCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCT 14600
Db      18153 CCCCGCCGTCGAGACCCTCGCCACCGACGAAGGCTTCACCCACTTCATCGAGGTCAGCGC 18212

Qy      14601 CACCACCCTCACCCACGACAACCTCCCCAACACCCCCACCACCACCCTCACCCTCACCCA 14660
Db      18213 CCACCCCGTCCTCACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGCACCCT---CCG 18269

Qy      14661 CCCCCACCACCACCCCCAAACCCACCTCCTCACCAACCTCGCCAAAACCACCACCACCTG 14720
Db      18270 CCGCGAACAGGGAGGCCAGGAGCGTCTGGTCACCTCACTCGCCGAAGCCTGGACCAACGG 18329

Qy      14721 GCACCCCCACCACTACACCCACCACCACAACCAACCCCACACCCACACCCACCTCGACCT 14780
Db      18330 CCTCACCATCGACTGGGCGCCCGTCC---TCCCCACCGCAACCGGCCACCACCCCGAGCT 18386

Qy      14781 CCCCACCTACCCCTTCCAACACCACCACTACTGGCTCCA 14819
Db      18387 CCCCACCTACGCCTTCCAGCGCCGTCACTACTGGCTCCA 18425
```

RESULT 3
US-09-836-821-5
; Sequence 5, Application US/09836821
; Publication No. US20030087405A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/836,821
; CURRENT FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA

; ORGANISM: Streptomyces venezuelae
US-09-836-821-5

Query Match 9.6%; Score 2951.8; DB 9; Length 36778;
Best Local Similarity 54.4%; Pred. No. 0;
Matches 8436; Conservative 0; Mismatches 5722; Indels 1361; Gaps 69;

```
Qy          8 GGATGGACGGCGGGGAAGAACCCCGCCCTGCGGCAGGGGAGGTCCTCGGAGTGGCCGACG 67
Db       3561 GGCCGGACGATCTGGCGGCAGCGCTGGCCGCGCCTGAGGGTCTGGTCCGGGGCGTGGCTT 3620

Qy         68 AGGCCGGACGGCGGCGTCGTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGG 127
Db       3621 CCGGTGTCGGGCGAGTGGCGTTCGTGTTCCCCGGGCAGGGCACGCAGTGGGCCGGCATGG 3680

Qy        128 GAAGGGAACTTCTCGACGCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCG 187
Db       3681 GTGCCGAACTGCTGGACTCTTCCGCGGTGTTCGCGGCGGCCATGGCCGAATGCGAGGCCG 3740

Qy        188 CGTTCGCGCCCTACGTCGACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTC 247
Db       3741 CACTCTCCCCGTACGTCGACTGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGC 3800

Qy        248 CCGGGCTGGACCGGGTGGACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGG 307
Db       3801 CCACGCTGGAGCGGGTCGATGTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGG 3860

Qy        308 CCGCCCTCTGGCGCTCGCAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCG 367
Db       3861 CTCGCGTGTGGCAGCACCACGGGGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCG 3920

Qy        368 AGATCGCGGCAGCCCACGTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGA 427
Db       3921 AGATCGCCGCCGCGTACGTCGCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGA 3980

Qy        428 CGCTTTGGAGCCAGGCACAGACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCG 484
Db       3981 CCCTGCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCG 4040

Qy        485 CCGCCACGCCGGATGAGCTCCTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGC 544
Db       4041 CGCTGAGCGAGGACGCCGTCCTGGAGCGACTGGCCGGGTTCGACG---------------- 4085

Qy        545 GGCTCGCCGTCGCAGCCGTCAACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGG 604
Db       4086 GGCTGTCCGTCGCCGCTGTGAACGGGCCCACCGCCACCGTGGTCTCCGGTGACCCCGTAC 4145

Qy        605 CCGTCGCGGACCTGGTGGCCGACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGG 664
Db       4146 AGATCGAAGAGCTTGCTCGGGCGTGTGAGGCCGATGGGGTCCGTGCGCGGGTCATTCCCG 4205

Qy        665 TGGACGTTCCCGCCCACTCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCC 724
Db       4206 TCGACTACGCGTCCCACAGCCGGCAGGTCGAGATCATCGAGAGCGAGCTCGCCGAGGTCC 4265

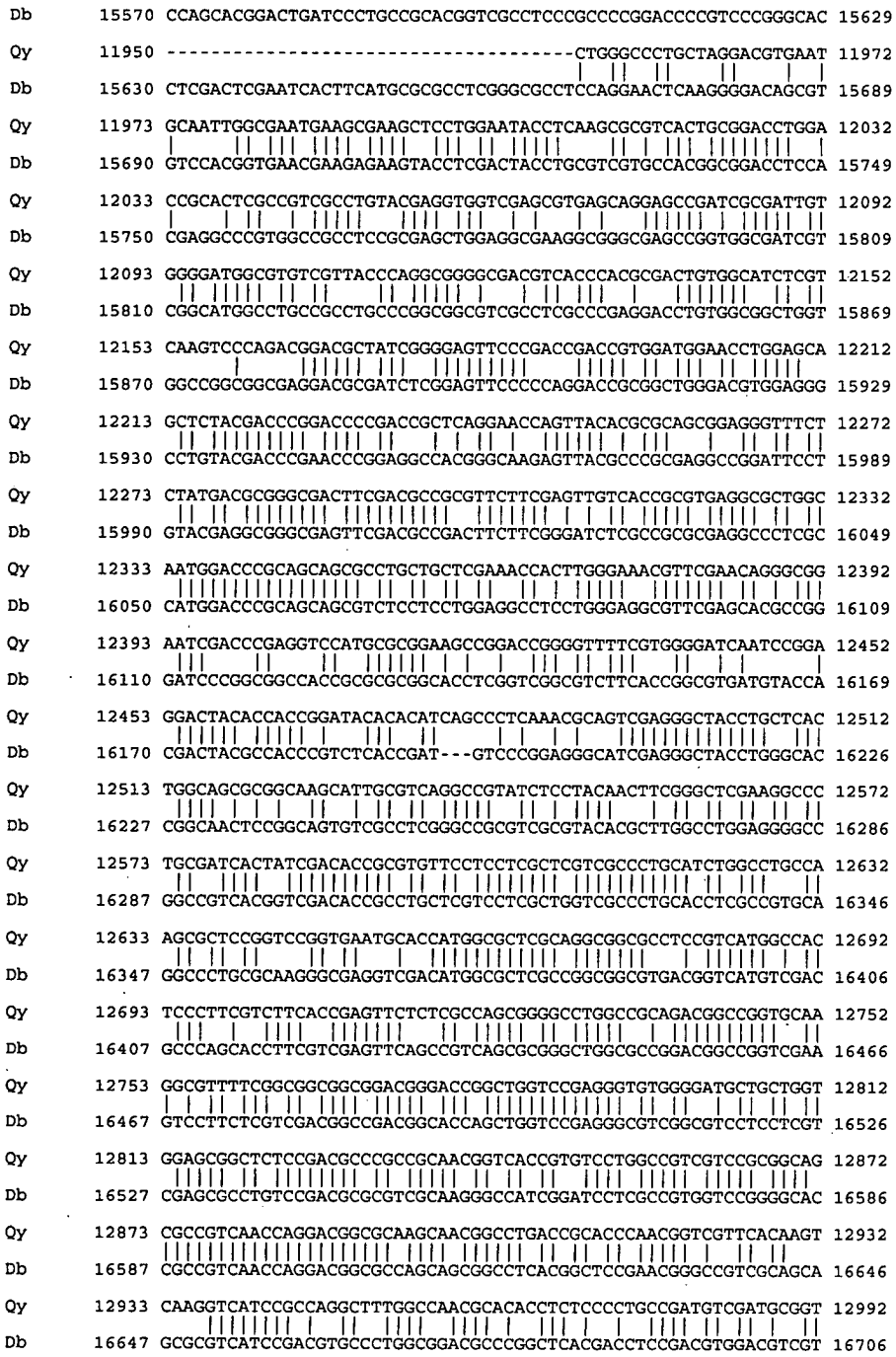
Qy        725 TGCTGCCCATCACCCCACGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCC 784
Db       4266 TCGCCGGGCTCAGCCCGCAGGCTCCGCGCGTGCCGTTCTTCTCGACACTCGAAGGCGCCT 4325

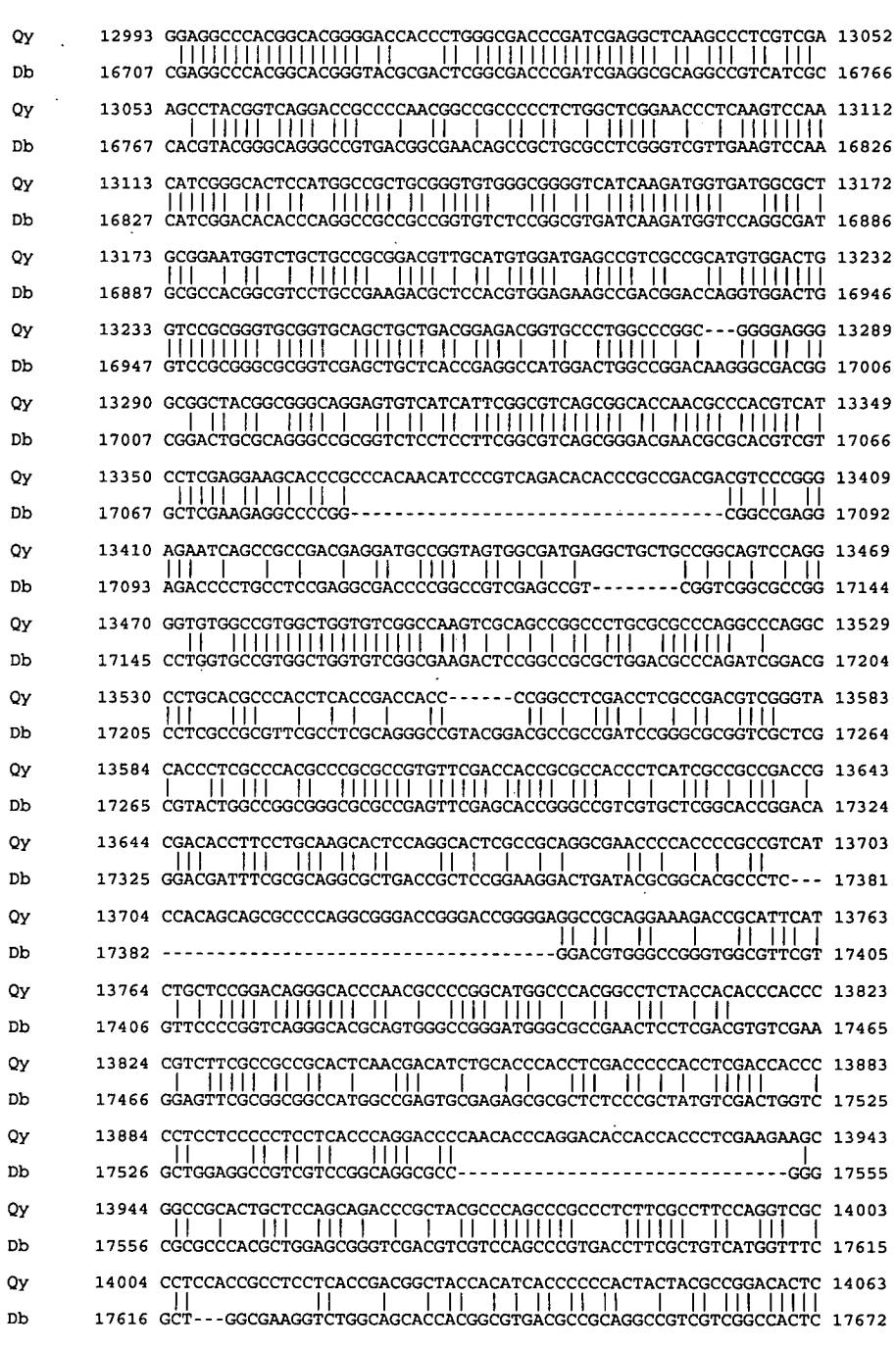
Qy        785 GCCTCGACACCCGCGAGCTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCC 844
Db       4326 GGATCACCGAGCCCGTGCTCGACGGCGGCTACTGGTACCGCAACCTGCGCCATCGTGTGG 4385

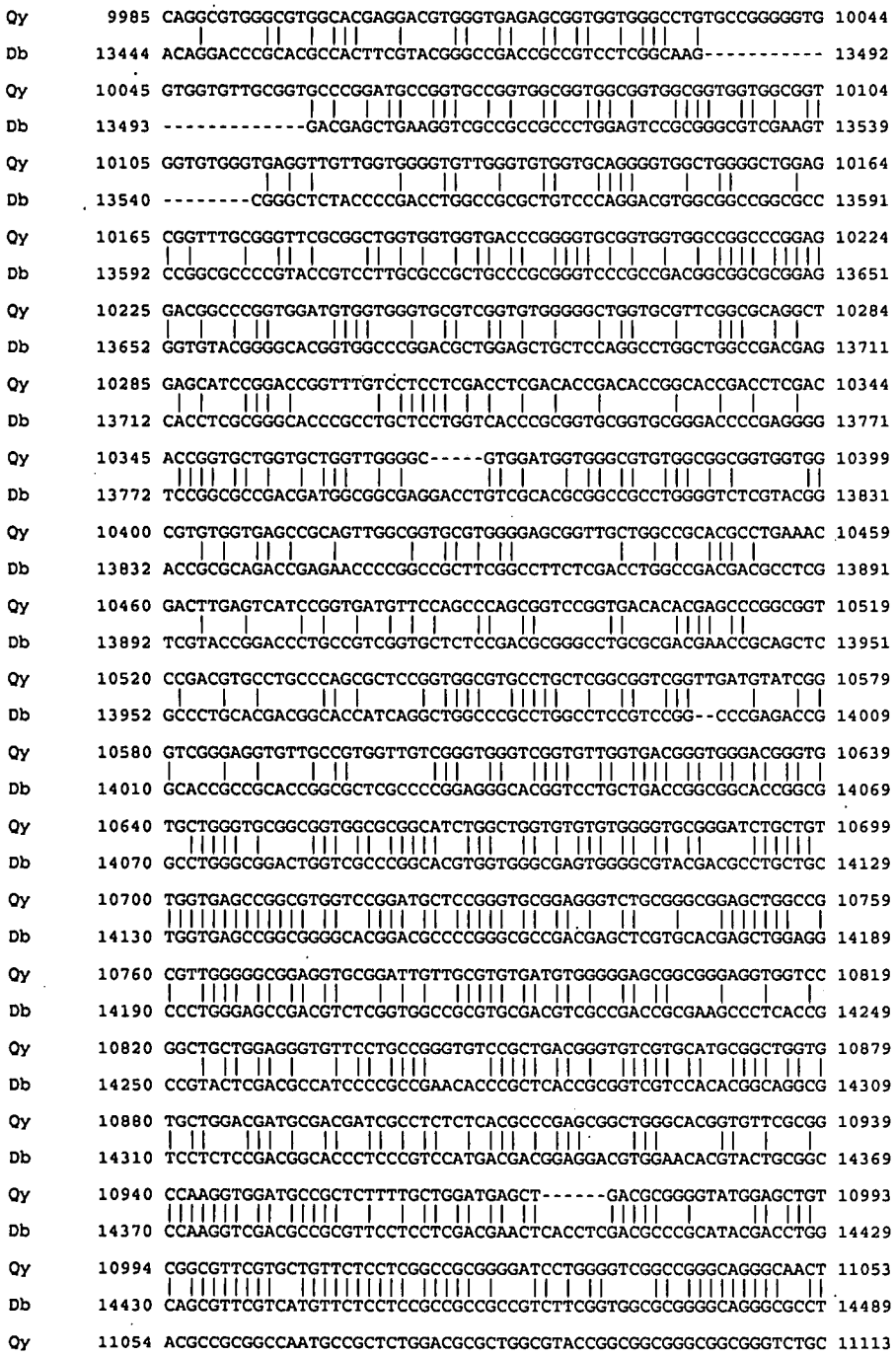
Qy        845 GGTTCGAGCCCGCCGCCCGG---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGA 901
Db       4386 GCTTCGCCCCGGCCGTCGAGACCCTGGCCACCGACGAGGGCTTCACCCACTTCGTCGAGG 4445

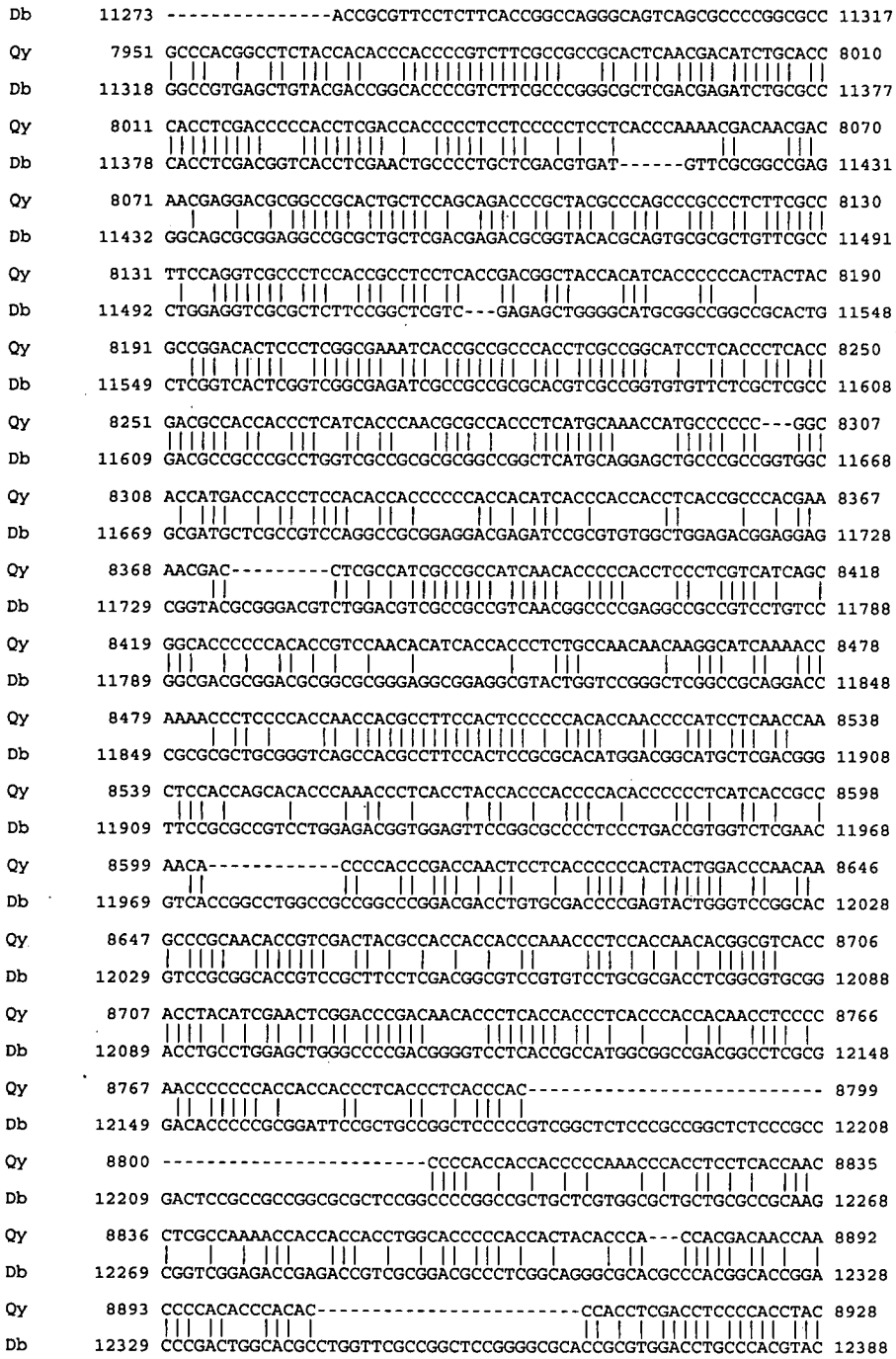
Qy        902 TGAGCCCGCACCCGGTGCTGACCATGGGCCTCCAGGAG---------------------- 939
Db       4446 TCAGCGCCCACCCCGTCCTCACCATGGCCCTCCCCGGGACCGTCACCGGTCTGGCGACCC 4505

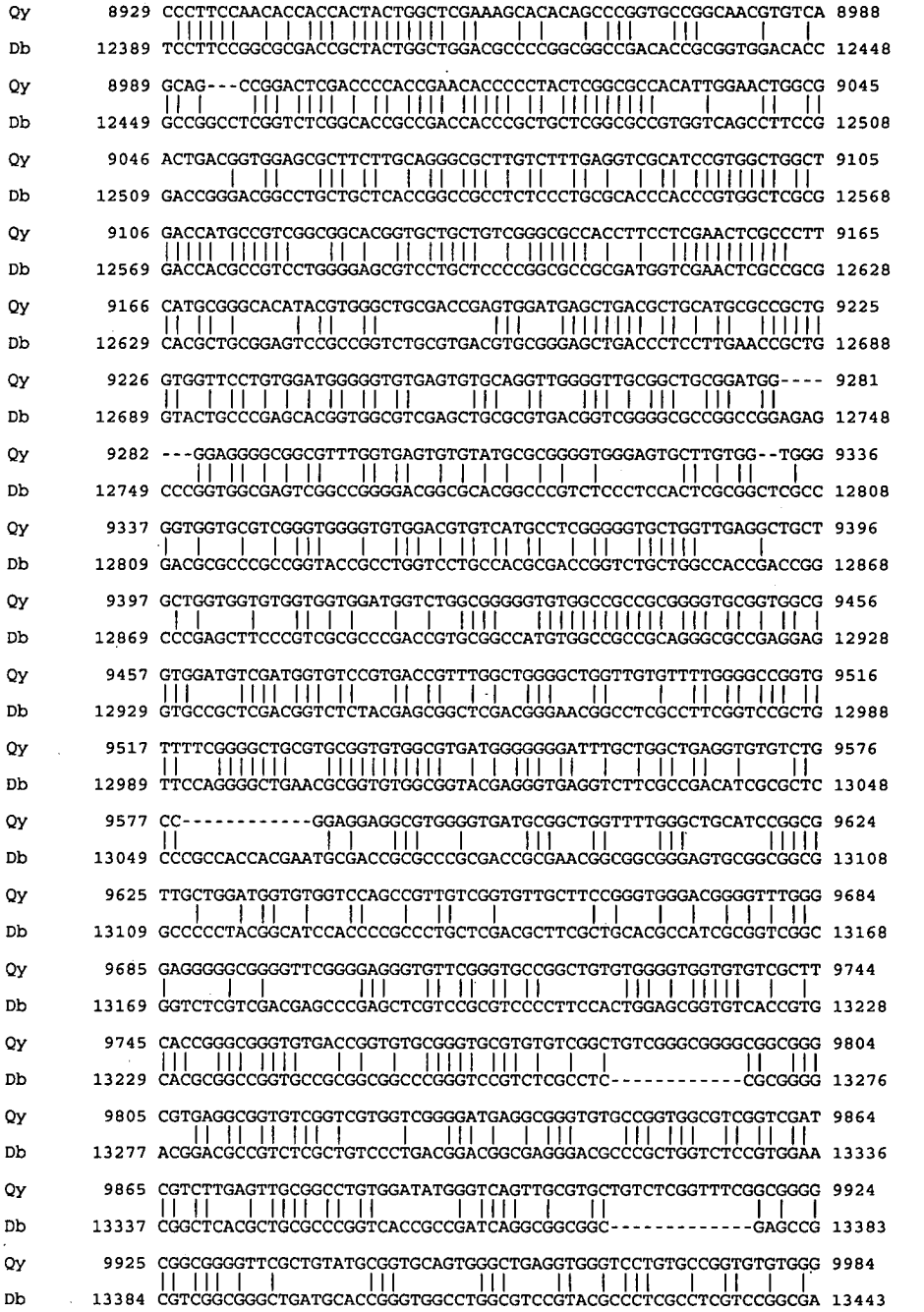
Qy        940 ----------------------------CTCGCCCCGGACCTGGGCGACACCACCGGCA 970
```

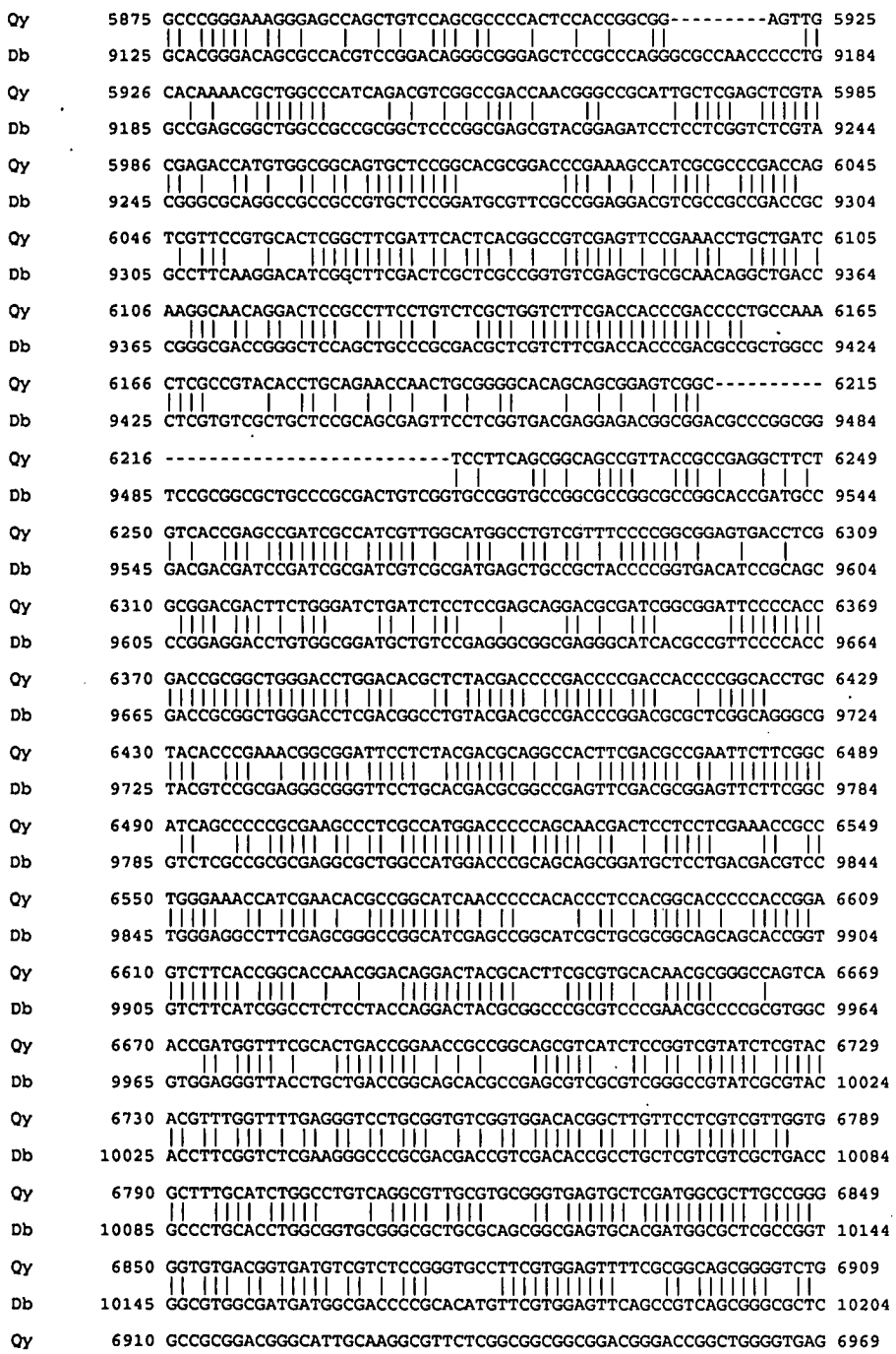


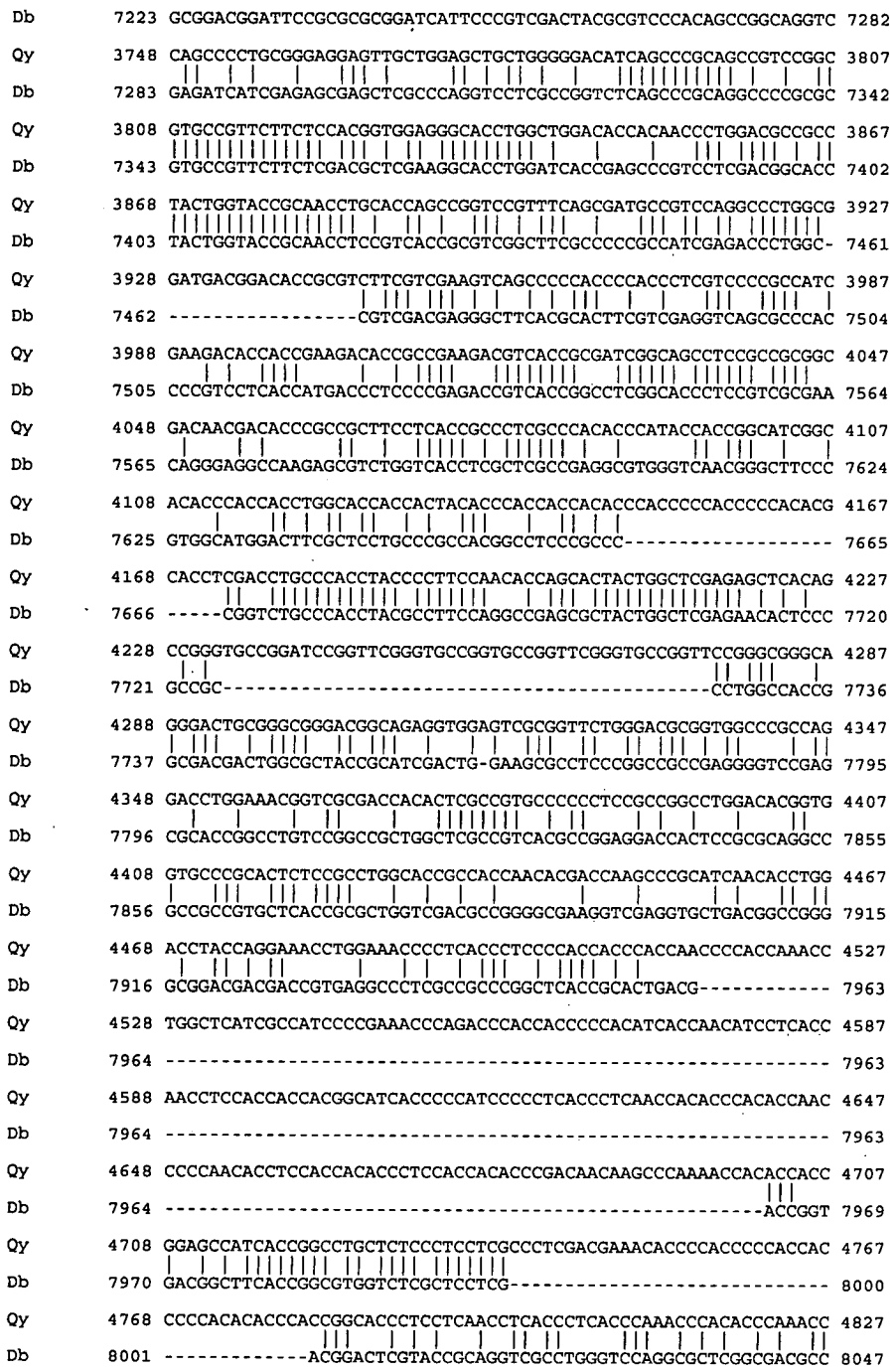


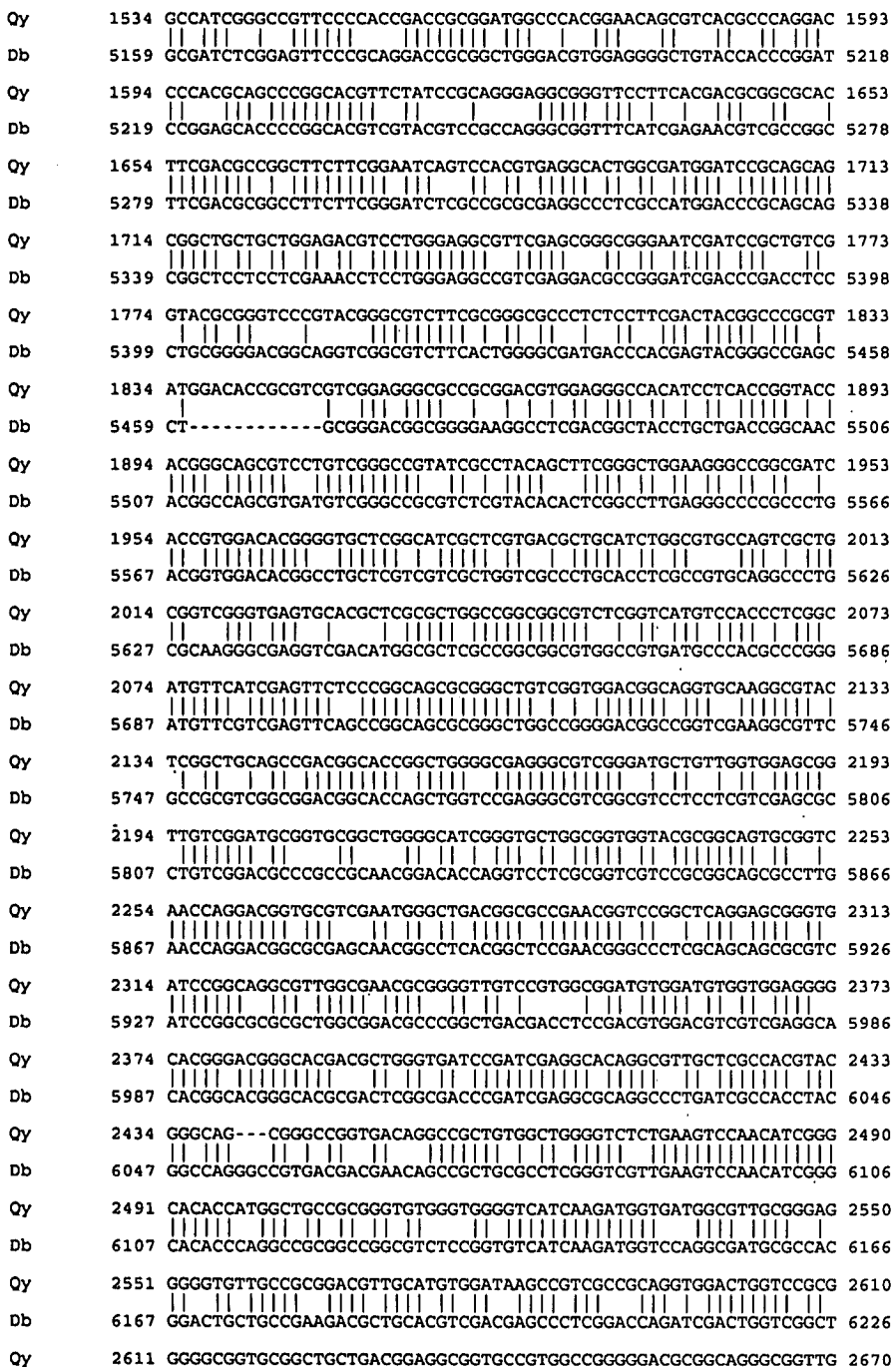

Qy 9985 CAGGCGTGGGCGTGGCACGAGGACGTGGGTGAGAGCGGTGGTGGGCCTGTGCCGGGGGTG 10044
Db 13444 ACAGGACCCGCACGCCACTTCGTACGGGCCGACCGCCGTCCTCGGCAAG---------- 13492
Qy 10045 GTGGTGTTGCGGTGCCCGGATGCCGGTGCCGGTGGCGGTGGCGGTGGCGGTGGTGGCGGT 10104
Db 13493 ----------GACGAGCTGAAGGTCGCCGCCGCCCTGGAGTCCGCGGGCGTCGAAGT 13539
Qy 10105 GGTGTGGGTGAGGTTGTTGGTGGGGTGTTGGGTGTGGTGCAGGGGTGGCTGGGGCTGGAG 10164
Db 13540 --------CGGGCTCTACCCCGACCTGGCCGCGCTGTCCCAGGACGTGGCGGCCGGCGCC 13591
Qy 10165 CGGTTTGCGGGTTCGCGGCTGGTGGTGGTGACCCGGGGTGCGGTGGTGGCCGGCCCGGAG 10224
Db 13592 CCGGCGCCCCGTACCGTCCTTGCGCCGCTGCCCGCGGGTCCCGCCGACGGCGGCGCGGAG 13651
Qy 10225 GACGGCCCGGTGGATGTGGTGGGTGCGTCGGTGTGGGGGCTGGTGCGTTCGGCGCAGGCT 10284
Db 13652 GGTGTACGGGGCACGGTGGCCCGGACGCTGGAGCTGCTCCAGGCCTGGCTGGCCGACGAG 13711
Qy 10285 GAGCATCCGGACCGGTTTGTCCTCCTCGACCTCGACACCGACACCGGCACCGACCTCGAC 10344
Db 13712 CACCTCGCGGGCACCCGCCTGCTCCTGGTCACCCGCGGTGCGGTGCGGGACCCCGAGGGG 13771
Qy 10345 ACCGGTGCTGGTGCTGGTTGGGGC-----GTGGATGGTGGGCGTGTGGCGGCGGTGGTGG 10399
Db 13772 TCCGGCGCCGACGATGGCGGCGAGGACCTGTCGCACGCGGCCGCCTGGGGTCTCGTACGG 13831
Qy 10400 CGTGTGGTGAGCCGCAGTTGGCGGTGCGTGGGGAGCGGTTGCTGGCCGCACGCCTGAAAC 10459
Db 13832 ACCGCGCAGACCGAGAACCCCGGCCGCTTCGGCCTTCTCGACCTGGCCGACGACGCCTCG 13891
Qy 10460 GACTTGAGTCATCCGGTGATGTTCCAGCCCAGCGGTCCGGTGACACACGAGCCCGGCGGT 10519
Db 13892 TCGTACCGGACCCTGCCGTCGGTGCTCTCCGACGCGGGCCTGCGCGACGAACCGCAGCTC 13951
Qy 10520 CCGACGTGCCTGCCCAGCGCTCCGGTGGCGTGCCTGCTCGGCGGTCGGTTGATGTATCGG 10579
Db 13952 GCCCTGCACGACGGCACCATCAGGCTGGCCCGCCTGGCCTCCGTCCGG--CCCGAGACCG 14009
Qy 10580 GTCGGGAGGTGTTGCCGTGGTTGTCGGGTGGGTCGGTGTTGGTGACGGGTGGGACGGGTG 10639
Db 14010 GCACCGCCGCACCGGCGCTCGCCCCGGAGGGCACGGTCCTGCTGACCGGCGGCACCGGCG 14069
Qy 10640 TGCTGGGTGCGGCGGTGGCGCGGCATCTGGCTGGTGTGTGTGGGGTGCGGGATCTGCTGT 10699
Db 14070 GCCTGGGCGGACTGGTCGCCCGGCACGTGGTGGGCGAGTGGGGCGTACGACGCCTGCTGC 14129
Qy 10700 TGGTGAGCCGGCGTGGTCCGGATGCTCCGGGTGCGGAGGGTCTGCGGGCGGAGCTGGCCG 10759
Db 14130 TGGTGAGCCGGCGGGGCACGGACGCCCCGGGCGCCGACGAGCTCGTGCACGAGCTGGAGG 14189
Qy 10760 CGTTGGGGGCGGAGGTGCGGATTGTTGCGTGTGATGTGGGGGAGCGGCGGGAGGTGGTCC 10819
Db 14190 CCCTGGGAGCCGACGTCTCGGTGGCCGCGTGCGACGTCGCCGACCGCGAAGCCCTCACCG 14249
Qy 10820 GGCTGCTGGAGGGTGTTCCTGCCGGGTGTCCGCTGACGGGTGTCGTGCATGCGGCTGGTG 10879
Db 14250 CCGTACTCGACGCCATCCCCGCCGAACACCCGCTCACCGCGGTCGTCCACACGGCAGGCG 14309
Qy 10880 TGCTGGACGATGCGACGATCGCCTCTCTCACGCCCGAGCGGCTGGGCACGGTGTTCGCGG 10939
Db 14310 TCCTCTCCGACGGCACCCTCCCGTCCATGACGACGGAGGACGTGGAACACGTACTGCGGC 14369
Qy 10940 CCAAGGTGGATGCCGCTCTTTTGCTGGATGAGCT------GACGCGGGGTATGGAGCTGT 10993
Db 14370 CCAAGGTCGACGCCGCGTTCCTCCTCGACGAACTCACCTCGACGCCCGCATACGACCTGG 14429
Qy 10994 CGGCGTTCGTGCTGTTCTCCTCGGCCGCGGGGATCCTGGGGTCGGCCGGGCAGGGCAACT 11053
Db 14430 CAGCGTTCGTCATGTTCTCCTCCGCCGCCGCCGTCTTCGGTGGCGCGGGGCAGGGCGCCT 14489
Qy 11054 ACGCCGCGGCCAATGCCGCTCTGGACGCGCTGGCGTACCGGCGGCGGGCGGCGGGTCTGC 11113
Db 14490 ACGCCGCCGCCAACGCCACCCTCGACGCCCTCGCCTGGCGCCGCCGGGCAGCCGGACTCC 14549
Qy 11114 CGGGGGTGTCGCTGGCGTGGGGGCTGTGGGAAGAGGCCAGCGGGATGACCGGGCACCTGG 11173
Db 14550 CCGCCCTCTCCCTCGGCTGGGGCCTCTGGGCCGAGACCAGCGGCATGACCGGCGAGCTCG 14609
Qy 11174 CCGGCACCGACCACCGGCGCATCATCCGTTCCGGTCTGCATCCCATGTCGACCCCGGACG 11233
Db 14610 GCCAGGCGGACCTGCGCCGGATGAGCCGCGCGGGCATCGGCGGGATCAGCGACGCCGAGG 14669
Qy 11234 CACTGGCCCTCTTCGATGCGGCCCTGGCTCTGGACCGGCCGGTCCTGCTGCCCGCCGACC 11293
Db 14670 GCATCGCGCTCCTCGACGCCGCCCTCCGCGACGACCGCCACCCGGTCCTGCTGCCCCTGC 14729
Qy 11294 TGCGTCCCGCCCCGCCCCTGCCG---------- 11316
Db 14730 GGCTCGACGCCGCCGGGCTGCGGGACGCGGCCGGGAACGACCCGGCCGGAATCCCGGCGC 14789
Qy 11317 ----------CCCCTGCTGCAGGACCTCCTGCCCGCCACCCGCCGCCGCACCACCCG 11363
Db 14790 TCTTCCGGGACGTCGTCGGCGCCAGGACCGTCCGGGCCCGGCCGTCCGCGGCCTCCGCCT 14849
Qy 11364 CACCACCACTACCGGTGGTGCGGACAACGGC---------- 11394
Db 14850 CGACGACAGCCGGGACGGCCGGCACGCCGGGGACGGCGGACGGCGCGGCGGAAACGGCGG 14909
Qy 11395 --GCCCAGCTGCACGCCCGGCTGGCCGGCCAGACACACGAACAACAGCACACCACCCTCC 11452
Db 14910 CGGTCACGCTCGCCGACCGGGCCGCCACCGTGGACGGGCCCGCACGGCAGCGCCTGCTGC 14969
Qy 11453 TCGCCCTGGTCCGCTCCCACATCGCCACCGTCCTGGGCCACACCACCCCCGACACCATCC 11512
Db 14970 TCGAGTTCGTCGTCGGCGAGGTCGCCGAAGTACTCGGCCACGCCCGCGGTCACCGGATCG 15029
Qy 11513 CCCCCGACCGCGCGTTCCGCGACCTCGGCTTCGACTCCCTCACCGCCGTCGAACTACGCA 11572
Db 15030 ACGCCGAACGGGGCTTCCTCGACCTCGGCTTCGACTCCCTGACCGCCGTCGAACTCCGCA 15089
Qy 11573 ACCGGCTCTCCCGCACCACCGGACTCCGCCTCCCCACCACCCTCGCCTTCGACCACCCCA 11632
Db 15090 ACCGGCTCAACTCCGCCGGTGGCCTCGCCCTCCCGGCGACCCTGGTCTTCGACCACCCAA 15149
Qy 11633 ACCCCACCACCCTCACCCACCACCTCCACACACAACTCC---------- 11671
Db 15150 GCCCGGCGGCACTCGCCTCCCACCTGGACGCCGAGCTGCCGCGCGGCGCCTCGGACCAGG 15209
Qy 11672 ----------AGCCACAACCGGACAACGCTGTCGC---------- 11696
Db 15210 ACGGAGCCGGGAACCGGAACGGGAACGAGAACGGGACGACGGCGTCCCGGAGCACCGCCG 15269
Qy 11697 -------CCCCGTGTTGGCGGAGCTCGACAAACTCGAATCCGCCCTCTCCGCCCTCGACA 11749
Db 15270 AGACGGACGCGCTGCTGGCACAACTGACCCGCCTGGAAGGCGCCTTGGTGCTGACGGGCC 15329
Qy 11750 AAACCGACAGCGCC---AGCGAAAGAGTCACCCTGCGGCTGAAGTCACTCATGTTGAGG- 11805
Db 15330 TCTCGGACGCCCCCGGGAGCGAAGAAGTCCTGGAGCACCTGCGGTCCCTGCGCTCGATGG 15389
Qy 11806 ----------TGGAACGCACCCCAGCATCCGACAGCCGAAAGCGCTGATG 11845
Db 15390 TCACGGGCGAGACCGGGACCGGGACCGCGTCCGGAGCCCCGGACGGCGCCGGGTCCGGCG 15449
Qy 11846 ACGACGA---------- 11852
Db 15450 CCGAGGACCGGCCCTGGGCGGCCGGGGACGGAGCCGGGGGCGGGAGTGAGGACGGCGCGG 15509
Qy 11853 -------GAAGTTCACATCGGCAACAGAGGCTGAGATTTTCAAATTCATTGACAACGACC 11905
Db 15510 GAGTGCCGGACTTCATGAACGCCTCGGCCGAGGAACTCTTCGGCCTCCTCGACCAGGACC 15569
Qy 11906 TCGGCCTGTCCTGAACCGGACGCCTGCCACTCCGCCCGTATCCG---------- 11949

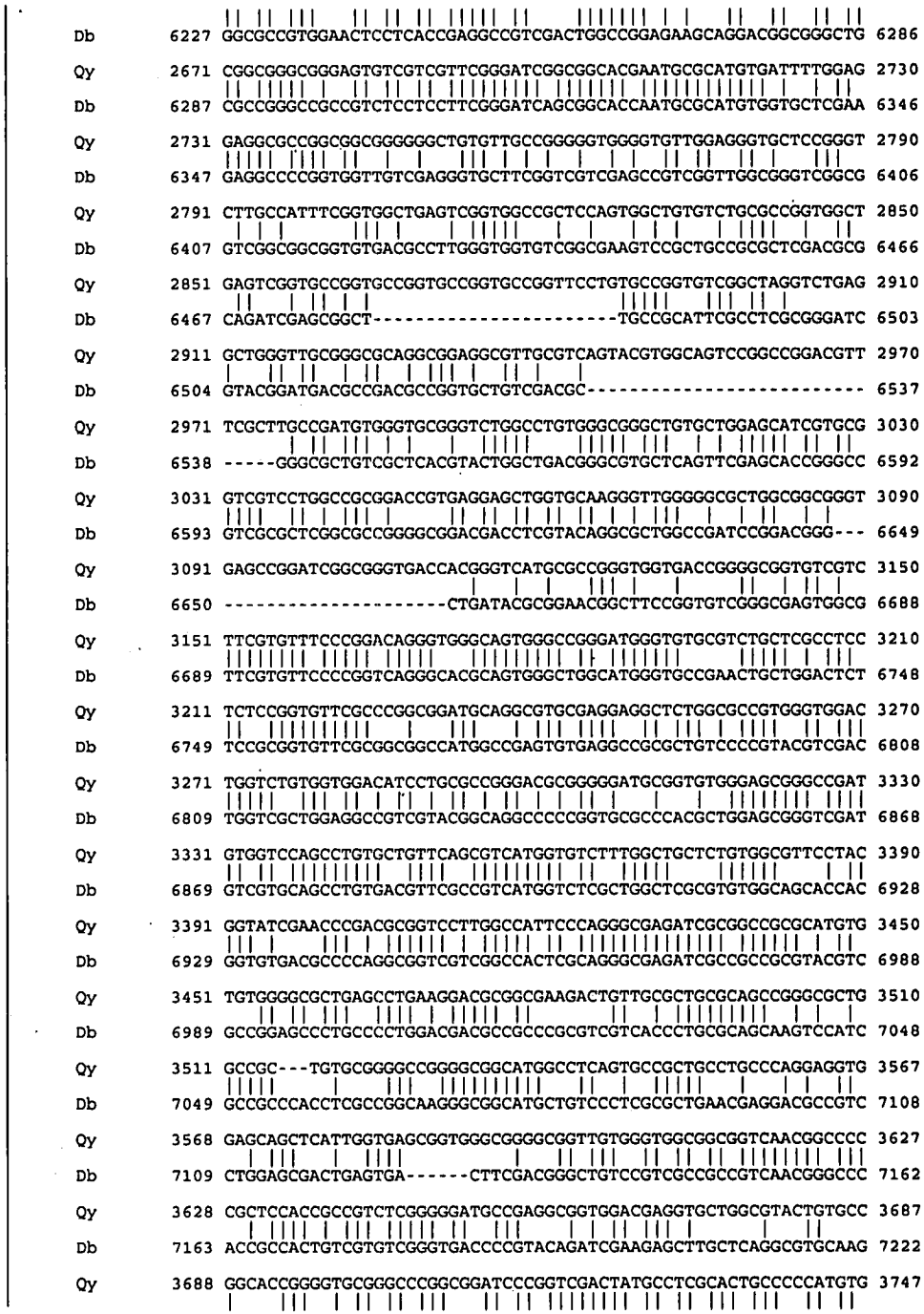





Qy 5875 GCCCGGGAAAGGGAGCCAGCTGTCCAGCGCCCCACTCCACCGGCGG---------AGTTG 5925
Db 9125 GCACGGGACAGCGCCACGTCCGGACAGGGCGGGAGCTCCGCCCAGGGCGCCAACCCCCTG 9184
Qy 5926 CACAAAACGCTGGCCCATCAGACGTCGGCCGACCAACGGGCCGCATTGCTCGAGCTCGTA 5985
Db 9185 GCCGAGCGGCTGGCCGCCCGCGGCTCCCGGCGAGCGTACGGAGATCCTCCTCGGTCTCGTA 9244
Qy 5986 CGAGACCATGTGGCGGCAGTGCTCCGGCACGCGGACCCGAAAGCCATCGCGCCCGACCAG 6045
Db 9245 CGGGCGCAGGCCGCCGCCGTGCTCCGGATGCGTTCGCCGGAGGACGTCGCCGCCGACCGC 9304
Qy 6046 TCGTTCCGTGCACTCGGCTTCGATTCACTCACGGCCGTCGAGTTCCGAAACCTGCTGATC 6105
Db 9305 GCCTTCAAGGACATCGGCTTCGACTCGCTCGCCGGTGTCGAGCTGCGCAACAGGCTGACC 9364
Qy 6106 AAGGCAACAGGACTCCGCCTTCCTGTCTCGCTGGTCTTCGACCACCCGACCCCTGCCAAA 6165
Db 9365 CGGGCGACCGGGCTCCAGCTGCCCGCGACGCTCGTCTTCGACCACCCGACGCCGCTGGCC 9424
Qy 6166 CTCGCCGTACACCTGCAGAACCAACTGCGGGGCACAGCAGCGGAGTCGGC---------- 6215
Db 9425 CTCGTGTCGCTGCTCCGCAGCGAGTTCCTCGGTGACGAGGAGACGGCGGACGCCCGGCGG 9484
Qy 6216 ----------TCCTTCAGCGGCAGCCGTTACCGCCGAGGCTTCT 6249
Db 9485 TCCGCGGCGCTGCCCGCGACTGTCGGTGCCGGTGCCGGCGCCGGCGCCGGCACCGATGCC 9544
Qy 6250 GTCACCGAGCCGATCGCCATCGTTGGCATGGCCTGTCGTTTCCCCGGCGGAGTGACCTCG 6309
Db 9545 GACGACGATCCGATCGCGATCGTCGCGATGAGCTGCCGCTACCCCGGTGACATCCGCAGC 9604
Qy 6310 GCGGACGACTTCTGGGATCTGATCTCCTCCGAGCAGGACGCGATCGGCGGATTCCCCACC 6369
Db 9605 CCGGAGGACCTGTGGCGGATGCTGTCCGAGGGCGGCGAGGGCATCACGCCGTTCCCCACC 9664
Qy 6370 GACCGCGGCTGGGACCTGGACACGCTCTACGACCCCGACCCCGACCACCCCGGCACCTGC 6429
Db 9665 GACCGCGGCTGGGACCTCGACGGCCTGTACGACGCCGACCCGGACGCGCTCGGCAGGGCG 9724
Qy 6430 TACACCCGAAACGGCGGATTCCTCTACGACGCAGGCCACTTCGACGCCGAATTCTTCGGC 6489
Db 9725 TACGTCCGCGAGGGCGGGTTCCTGCACGACGCGGCCGAGTTCGACGCGGAGTTCTTCGGC 9784
Qy 6490 ATCAGCCCCCGCGAAGCCCTCGCCATGGACCCCCAGCAACGACTCCTCCTCGAAACCGCC 6549
Db 9785 GTCTCGCCGCGCGAGGCGCTGGCCATGGACCCGCAGCAGCGGATGCTCCTGACGACGTCC 9844
Qy 6550 TGGGAAACCATCGAACACGCCGGCATCAACCCCCACACCCTCCACGGCACCCCCACCGGA 6609
Db 9845 TGGGAGGCCTTCGAGCGGGCCGGCATCGAGCCGGCATCGCTGCGCGGCAGCAGCACCGGT 9904
Qy 6610 GTCTTCACCGGCACCAACGGACAGGACTACGCACTTCGCGTGCACAACGCGGGCCAGTCA 6669
Db 9905 GTCTTCATCGGCCTCTCCTACCAGGACTACGCGGCCCGCGTCCCGAACGCCCCGCGTGGC 9964
Qy 6670 ACCGATGGTTTCGCACTGACCGGAACCGCCGGCAGCGTCATCTCCGGTCGTATCTCGTAC 6729
Db 9965 GTGGAGGGTTACCTGCTGACCGGCAGCACGCCGAGCGTCGCGTCGGGCCGTATCGCGTAC 10024
Qy 6730 ACGTTTGGTTTTGAGGGTCCTGCGGTGTCGGTGGACACGGCTTGTTCCTCGTCGTTGGTG 6789
Db 10025 ACCTTCGGTCTCGAAGGGCCCGCGACGACCGTCGACACCGCCTGCTCGTCGTCGCTGACC 10084
Qy 6790 GCTTTGCATCTGGCCTGTCAGGCGTTGCGTGCGGGTGAGTGCTCGATGGCGCTTGCCGGG 6849
Db 10085 GCCCTGCACCTGGCGGTGCGGGCGCTGCGCAGCGGCGAGTGCACGATGGCGCTCGCCGGT 10144
Qy 6850 GGTGTGACGGTGATGTCGTCTCCGGGTGCCTTCGTGGAGTTTTCGCGGCAGCGGGGTCTG 6909
Db 10145 GGCGTGGCGATGATGGCGACCCCGCACATGTTCGTGGAGTTCAGCCGTCAGCGGGCGCTC 10204
Qy 6910 GCCGCGGACGGGCATTGCAAGGCGTTCTCGGCGGCGGCGGACGGGACCGGCTGGGGTGAG 6969
Db 10205 GCCCCGGACGGCCGCAGCAAGGCCTTCTCGGCGGACGCCGACGGGTTCGGCGCCGCGGAG 10264
Qy 6970 GGTGTGGGGATGCTGCTGGTGGAGCGGCTCTCCGACGCCCATCGCAACGGTCACCGTGTC 7029
Db 10265 GGCGTCGGCCTGCTGCTCGTGGAGCGGCTCTCGGACGCGCGGCGCAACGGTCACCCGGTG 10324
Qy 7030 CTGGCCGTGGTGCGTGGCAGTGCGGTCAACCAGGACGGTGCGAGCAACGGTCTGACCGCG 7089
Db 10325 CTCGCCGTGGTCCGCGGTACCGCCGTCAACCAGGACGGCGCCAGCAACGGGCTGACCGCG 10384
Qy 7090 CCCAACGGGCCGTCCCAGCAGCGTGTCATCCGCCAGGCCCTCGCCAACGCCGGCTTGTCG 7149
Db 10385 CCCAACGGACCCTCGCAGCAGCGGGTGATCCGGCAGGCGCTCGCCGACGCCCGGCTGGCA 10444
Qy 7150 GCCGGTGATGTCGACGCGGTGGAGGCCCACGGCACCGGCACCACTTTGGGCGACCCGATC 7209
Db 10445 CCCGGCGACATCGACGCCGTCGAGACGCACGGCACGGGAACCTCGCTGGGCGACCCCATC 10504
Qy 7210 GAGGCCCAGGCCCTCCTCGCGACCTACGGACAGGACCGTGCCGGCGAGGGGCCGCTGTGG 7269
Db 10505 GAGGCCCAGGGCCTCCAGGCCACGTACGGCAAGGAGCGGCCCGCGGAACGGCCGCTCGCC 10564
Qy 7270 CTGGGCTCGGTCAAGTCCAATGTCGGTCACACACAGGCTGCCGCGGGCGTCGCCGGGGTG 7329
Db 10565 ATCGGCTCCGTGAAGTCCAACATCGGACACACCCAGGCCGCGGCCGGTGCGGCGGGCATC 10624
Qy 7330 ATCAAGATGGTGATGGCGCTGCGGCATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAG 7389
Db 10625 ATCAAGATGGTCCTCGCGATGCGCCACGGCACCCTGCCGAAGACCCTCCACGCCGACGAG 10684
Qy 7390 CCGTCGCCGCATGTGGACTGGTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCC 7449
Db 10685 CCGAGCCCGCACGTCGACTGGGCGAACAGCGGCCTGGCCCTCGTCACCGAGCCGATCGAC 10744
Qy 7450 TGGCCCGGCGGGGAGGGGCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGC 7509
Db 10745 TGGCCGGCCGGCACCGGTCCG---CGCCGCGCCGCCGTCTCCTCCTTCGGCATCAGCGGG 10801
Qy 7510 ACCAACGCCCACGTCATCCTCGAAGAAGCACCCGCCGACGACGTT--------------- 7554
Db 10802 ACGAACGCGCACGTCGTGCTGGAGCAGGCGCCGGATGCTGCTGGTGAGGTGCTTGGGGCC 10861
Qy 7555 ----------CCGGGGGGACCACCC 7569
Db 10862 GATGAGGTGCCTGAGGTGTCTGAGACGGTAGCGATGGCTGGGACGGCTGGGACCTCCGAG 10921
Qy 7570 GCCGGCGAGGGTGACGCGGGCAGCGACGATGAGGCTGCTGCCGGCAGTCCTGGGGTGT-- 7627
Db 10922 GTCGCTGAGGGCTCTGAGGCCTCCGAGGCCCCCGCGGCCCCCGGCAGCCGTGAGGCGTCC 10981
Qy 7628 ------------GGCCGTGGCTGGTGTCGGCCAAGTCGCAGCCGGCCCTGCGCGCCCAG 7674
Db 10982 CTCCCCGGGCACCTGCCCTGGGTGCTGTCCGCCAAGGACGAGCAGTCGCTGCGCGGCCAG 11041
Qy 7675 GCCCAGGCCCTGCACGCCCACCTCACCGACCACCCCGGCCTCGA--------------- 7718
Db 11042 GCCGCCGCCCTGCACGCGTGGCTGTCCGAGCCCGCCGCCGACCTGTCGGACGCGGACGGA 11101
Qy 7719 --------CCTCGCGGATGTCGGATACACCCTCGCCCACGCCCGCGCCGTGTTCGACCAC 7770
Db 11102 CCGGCCCGCCTGCGGGACGTCGGGTACACGCTCGCCACGAGCCGTACCGCCTTCGCGCAC 11161
Qy 7771 CGCGCCACCCTCATCGCCGCGGACCGCGACACGTTCCTGCAAGCACTCCAGGCACTCGCC 7830
Db 11162 CGCGCCGCCGTGACCGCCGCCGACCGGGACGGGTTCCTGGACGGGCTGGCCACGCTGGCC 11221
Qy 7831 GCAGGCGAGCCCCACCCCGCCGTCATCCACAGCAGCGCCCCGGGCGGGACCGGGACCGGG 7890
Db 11222 CAGGGCGGCACCTCGGCCCACGTCCACCTGGACACCGCCCGGGACGGCACC--------- 11272
Qy 7891 GAGGCCGCAGGAAAGACCGCATTCATCTGCTCCGGACAGGGCACCCAACGCCCCGGCATG 7950


Db 7223 GCGGACGGATTCCGCGCGCGGATCATTCCCGTCGACTACGCGTCCCACAGCCGGCAGGTC 7282
Qy 3748 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 3807
Db 7283 GAGATCATCGAGAGCGAGCTCGCCCAGGTCCTCGCCGGTCTCAGCCCGCAGGCCCCGCGC 7342
Qy 3808 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 3867
Db 7343 GTGCCGTTCTTCTCGACGCTCGAAGGCACCTGGATCACCGAGCCCGTCCTCGACGGCACC 7402
Qy 3868 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 3927
Db 7403 TACTGGTACCGCAACCTCCGTCACCGCGTCGGCTTCGCCCCCGCCATCGAGACCCTGGC- 7461
Qy 3928 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 3987
Db 7462 ----------CGTCGACGAGGGCTTCACGCACTTCGTCGAGGTCAGCGCCCAC 7504
Qy 3988 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 4047
Db 7505 CCCGTCCTCACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGCACCCTCCGTCGCGAA 7564
Qy 4048 GACAACGACACCCGCCGCTTCCTCACCGCCCTCGCCCACACCCATACCACCGGCATCGGC 4107
Db 7565 CAGGGAGGCCAAGAGCGTCTGGTCACCTCGCTCGCCGAGGCGTGGGTCAACGGGCTTCCC 7624
Qy 4108 ACACCCACCACCTGGCACCACCACTACACCCACCACCACACCCACCCCCACCCCCACACG 4167
Db 7625 GTGGCATGGACTTCGCTCCTGCCCGCCACGGCCTCCCGCCC---------- 7665
Qy 4168 CACCTCGACCTGCCCACCTACCCCTTCCAACACCAGCACTACTGGCTCGAGAGCTCACAG 4227
Db 7666 -----CGGTCTGCCCACCTACGCCTTCCAGGCCGAGCGCTACTGGCTCGAGAACACTCCC 7720
Qy 4228 CCGGGTGCCGGATCCGGTTCGGGTGCCGGTGCCGGTTCGGGTGCCGGTTCCGGGCGGGCA 4287
Db 7721 GCCGC----------CCTGGCCACCG 7736
Qy 4288 GGGACTGCGGGCGGGACGGCAGAGGTGGAGTCGCGGTTCTGGGACGCGGTGGCCCGCCAG 4347
Db 7737 GCGACGACTGGCGCTACCGCATCGACTG-GAAGCGCCTCCCGGCCGCCGAGGGGTCCGAG 7795
Qy 4348 GACCTGGAAACGGTCGCGACCACACTCGCCGTGCCCCCCTCCGCCGGCCTGGACACGGTG 4407
Db 7796 CGCACCGGCCTGTCCGGCCGCTGGCTCGCCGTCACGCCGGAGGACCACTCCGCGCAGGCC 7855
Qy 4408 GTGCCCGCACTCTCCGCCTGGCACCGCCACCAACACGACCAAGCCCGCATCAACACCTGG 4467
Db 7856 GCCGCCGTGCTCACCGCGCTGGTCGACGCCGGGGCGAAGGTCGAGGTGCTGACGGCCGGG 7915
Qy 4468 ACCTACCAGGAAACCTGGAAACCCCTCACCCTCCCCACCACCCACCAACCCCACCAAACC 4527
Db 7916 GCGGACGACGACCGTGAGGCCCTCGCCGCCCGGCTCACCGCACTGACG---------- 7963
Qy 4528 TGGCTCATCGCCATCCCCGAAACCCAGACCCACCACCCCCACATCACCAACATCCTCACC 4587
Db 7964 ---------- 7963
Qy 4588 AACCTCCACCACCACGGCATCACCCCCATCCCCCTCACCCTCAACCACACCCACACCAAC 4647
Db 7964 ---------- 7963
Qy 4648 CCCCAACACCTCCACCACACCCTCCACCACACCCGACAACAAGCCCAAAACCACACCACC 4707
Db 7964 ----------ACCGGT 7969
Qy 4708 GGAGCCATCACCGGCCTGCTCTCCCTCCTCGCCCTCGACGAAACACCCCACCCCCACCAC 4767
Db 7970 GACGGCTTCACCGGCGTGGTCTCGCTCCTCG---------- 8000
Qy 4768 CCCCACACACCCACCGGCACCCTCCTCAACCTCACCCTCACCCAAACCCACACCCAAACC 4827
Db 8001 ----------ACGGACTCGTACCGCAGGTCGCCTGGGTCCAGGCGCTCGGCGACGCC 8047
Qy 4828 CACCCACCAACCCCCCTCTGGTACGCCACCACCAACGCCACCACCACCCACCCCAACGAC 4887
Db 8048 GGAATCAAGGCGCCCCTGTGGTCCGTCACCCAGGGCGCGGTCTCCGTCGGACGTCTCGAC 8107
Qy 4888 CCCCTCACACACCCCACCCAAGCCCAAACCTGGGGACTCGCCCGCACCACCCTCCTCGAA 4947
Db 8108 ACCCCCGCCGACCCCGACCGGGCCATGCTCTGGGGCCTCGGCCGCGTCGTCGCCCTTGAG 8167
Qy 4948 CACCCCACCCACACCGCCGGAATCATCGACCTCCCCACCACCCCCACCCCCCACACCCTC 5007
Db 8168 CACCCCGAACGCTGGGCCGGCCTCGTCGACCTCCCCGCCCAGCCCGATGCCGCCGCCCTC 8227
Qy 5008 CAGCACCTCACCCAAACCCTCACCCAACCCCACCACCAAACCCAACTCGCCATCCGCACC 5067
Db 8228 GCCCACCTCGTCACCGCACTCTCCGGCGCCACCGGCGAGGACCAGATCGCCATCCGCACC 8287
Qy 5068 ACCGGCACCCACACCCGCCGCCTCACCCCCACCACCCTCACCCCCACACACCAACCACCC 5127
Db 8288 ACCGGACTCCACGCCCGCCGCCTCGCCCGCGCACCCCTCCACGGACGTCGGCCCACCCGC 8347
Qy 5128 ACCCCCACCCCCCACGGAACCACCCTCATCACCGGCGGAACCGGCGCCCTCGCCACCCAC 5187
Db 8348 GACTGGCAGCCCCACGGCACCGTCCTCATCACCGGCGGCACCGGAGCCCTCGGCAGCCA- 8406
Qy 5188 CTCACCCACCACCTCACCACCCACCAACCCACCCAACACCTCCTCCTCACCAGCCGAACC 5247
Db 8407 --CGCCGCACGCTGGATGGCCCACCACGGAGCCGAACACCTCCTCCTCGTCAGCCGCAGC 8464
Qy 5248 GGCCCCCACACCCCCCACGCACAACACCTCACCACCCAACTCCAACAAAAAGGCATCCAC 5307
Db 8465 GGCGAACAAGCCCCCGGAGCCACCCAACTCACCGCCGAACTCACCGCATCGGGCGCCCGC 8524
Qy 5308 CTCACCATCACCACCTGCGACACCAGCAACCCAGACCAACTCCAACAACTCCTCAACACC 5367
Db 8525 GTCACCATCGCCGCCTGCGACGTCGCCGACCCCCACGCCATGCGCACCCTCCTCGACGCC 8584
Qy 5368 ATCCCCCCCACAACACCCCCTCACCACCGTCATCCACACCGCAGGCATCCTCGACGACGCC 5427
Db 8585 ATCCCCGCCGAGACGCCCCTCACCGCCGTCGTCCACACCGCCGGCGCGCTCGACGACGGC 8644
Qy 5428 ACCCTCACCAACCTCACCCCCACCCAACTCAACAACGTCCTCCGCGCCAAAGCCCACAGC 5487
Db 8645 ATCGTGGACACGCTGACCGCCGAGCAGGTCCGGCGGGCCCACCGTGCGAAGGCCGTCGGC 8704
Qy 5488 GCCCACCTCCTCCACCAACTCACCCAACACACCCCCCTCACCGCCTTCGTCCTCTACTCC 5547
Db 8705 GCCTCGGTGCTCGACGAGCTGACCCGGGACCTCGACCTCGACGCGTTCGTGCTCTTCTCG 8764
Qy 5548 TCCGCCGCCGCCACCTTCGGCGCACCCGGCCAAGCCAACTACGCCGCAGCCAACGCCTAC 5607
Db 8765 TCCGTGTCGAGCACTCTGGGCATCCCCGGTCAGGGCAACTACGCCCCGCACAACGCCTAC 8824
Qy 5608 CTCGACGCCCTCGCCCACCACCGCCACACCCACCACCTCCCCGCCACCAGCATCGCCTGG 5667
Db 8825 CTCGACGCCCTCGCGGCTCGCCGCCGGGCCACCGGCCGGTCCGCCGTCTCGGTGGCCTGG 8884
Qy 5668 GGCACCTGGCAAGGAAACGGACTCGCTGATTCGGACAAGGCCCGCGCATATCTCGACCGC 5727
Db 8885 GGACCGTGGGACGGTGGCGGCATGGCCGCCGGTGACGGCGTGGCCGAGCGGCTGCGCAAC 8944
Qy 5728 CGCGGGTTTCGACCCATGTCACCCGAGTTGGCCACGGCAGC---------- 5768
Db 8945 CACGGCGTGCCCGGCATGGACCCGGAACTCGCCCTGGCCGCACTGGAGTCCGCGCTCGGC 9004
Qy 5769 --------GGTCACGCAGGCGATCGCGGACACCGAACGGCCGTATGTCGTCATCGCCGAC 5820
Db 9005 CGGGACGAGACCGCGATCACCGTCGCGGACATCGACTGGGACCGCTTCTACCTCGCGTAC 9064
Qy 5821 ATCGACTGGAGCAAGATCGAACACACCTCTCAGACCAGCGACCTG------GTGAGCGCG 5874
Db 9065 TCCTCCGGTCGCCCGCAGCCCCTCGTCGAGGAGCTGCCCGAGGTGCGGCGCATCATCGAC 9124


Qy 1534 GCCATCGGGCCGTTCCCCACCGACCGCGGATGGCCCACGGAACAGCGTCACGCCCAGGAC 1593
Db 5159 GCGATCTCGGAGTTCCCGCAGGACCGCGGCTGGGACGTGGAGGGGCTGTACCACCCGGAT 5218
Qy 1594 CCCACGCAGCCCGGCACGTTCTATCCGCAGGGAGGCGGGTTCCTTCACGACGCGGCGCAC 1653
Db 5219 CCGGAGCACCCCGGCACGTCGTACGTCCGCCAGGGCGGTTTCATCGAGAACGTCGCCGGC 5278
Qy 1654 TTCGACGCCGGCTTCTTCGGAATCAGTCCACGTGAGGCACTGGCGATGGATCCGCAGCAG 1713
Db 5279 TTCGACGCGGCCTTCTTCGGGATCTCGCCGCGCGAGGCCCTCGCCATGGACCCGCAGCAG 5338
Qy 1714 CGGCTGCTGCTGGAGACGTCCTGGGAGGCGTTCGAGCGGGCGGGAATCGATCCGCTGTCG 1773
Db 5339 CGGCTCCTCCTCGAAACCTCCTGGGAGGCCGTCGAGGACGCCGGGATCGACCCGACCTCC 5398
Qy 1774 GTACGCGGGTCCCGTACGGGCGTCTTCGCGGGCGCCCTCTCCTTCGACTACGGCCCGCGT 1833
Db 5399 CTGCGGGGACGGCAGGTCGGCGTCTTCACTGGGGCGATGACCCACGAGTACGGGCCGAGC 5458
Qy 1834 ATGGACACCGCGTCGTCGGAGGGCGCCGCGGACGTGGAGGGCCACATCCTCACCGGTACC 1893
Db 5459 CT----------GCGGGACGGCGGGGAAGGCCTCGACGGCTACCTGCTGACCGGCAAC 5506
Qy 1894 ACGGGCAGCGTCCTGTCGGGCCGTATCGCCTACAGCTTCGGGCTGGAAGGGCCGGCGATC 1953
Db 5507 ACGGCCAGCGTGATGTCGGGCCGCGTCTCGTACACACTCGGCCTTGAGGGCCCCGCCCTG 5566
Qy 1954 ACCGTGGACACGGGGTGCTCGGCATCGCTCGTGACGCTGCATCTGGCGTGCCAGTCGCTG 2013
Db 5567 ACGGTGGACACGGCCTGCTCGTCGTCGCTGGTCGCCCTGCACCTCGCCGTGCAGGCCCTG 5626
Qy 2014 CGGTCGGGTGAGTGCACGCTCGCGCTGGCCGGCGGCGTCTCGGTCATGTCCACCCTCGGC 2073
Db 5627 CGCAAGGGCGAGGTCGACATGGCGCTCGCCGGCGGCGTGGCCGTGATGCCCACGCCCGGG 5686
Qy 2074 ATGTTCATCGAGTTCTCCCGGCAGCGCGGGCTGTCGGTGGACGGCAGGTGCAAGGCGTAC 2133
Db 5687 ATGTTCGTCGAGTTCAGCCGGCAGCGCGGGCTGGCCGGGGACGGCCGGTCGAAGGCGTTC 5746
Qy 2134 TCGGCTGCAGCCGACGGCACCGGCTGGGGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGG 2193
Db 5747 GCCGCGTCGGCGGACGGCACCAGCTGGTCCGAGGGCGTCGGCGTCCTCCTCGTCGAGCGC 5806
Qy 2194 TTGTCGGATGCGGTGCGGCTGGGGCATCGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTC 2253
Db 5807 CTGTCGGACGCCCGCCGCAACGGACACCAGGTCCTCGCGGTCGTCCGCGGCAGCGCCTTG 5866
Qy 2254 AACCAGGACGGTGCGTCGAATGGGCTGACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTG 2313
Db 5867 AACCAGGACGGCGCGAGCAACGGCCTCACGGCTCCGAACGGGCCCTCGCAGCAGCGCGTC 5926
Qy 2314 ATCCGGCAGGCGTTGGCGAACGCGGGGTTGTCCGTGGCGGATGTGGATGTGGTGGAGGGG 2373
Db 5927 ATCCGGCGCGCGCTGGCGGACGCCCGGCTGACGACCTCCGACGTGGACGTCGTCGAGGCA 5986
Qy 2374 CACGGGACGGGCACGACGCTGGGTGATCCGATCGAGGCACAGGCGTTGCTCGCCACGTAC 2433
Db 5987 CACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCGCAGGCCCTGATCGCCACCTAC 6046
Qy 2434 GGGCAG---CGGGCCGGTGACAGGCCGCTGTGGCTGGGGTCTCTGAAGTCCAACATCGGG 2490
Db 6047 GGCCAGGGCCGTGACGACGAACAGCCGCTGCGCCTCGGGTCGTTGAAGTCCAACATCGGG 6106
Qy 2491 CACACCATGGCTGCCGCGGGTGTGGGTGGGGTCATCAAGATGGTGATGGCGTTGCGGGAG 2550
Db 6107 CACACCCAGGCCGCGGCCGGCGTCTCCGGTGTCATCAAGATGGTCCAGGCGATGCGCCAC 6166
Qy 2551 GGGGTGTTGCCGCGGACGTTGCATGTGGATAAGCCGTCGCCGCAGGTGGACTGGTCCGCG 2610
Db 6167 GGACTGCTGCCGAAGACGCTGCACGTCGACGAGCCCTCGGACCAGATCGACTGGTCGGCT 6226
Qy 2611 GGGGCGGTGCGGCTGCTGACGGAGGCGGTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTG 2670


Db 6227 GGCGCCGTGGAACTCCTCACCGAGGCCGTCGACTGGCCGGAGAAGCAGGACGGCGGGCTG 6286
Qy 2671 CGGCGGGCGGGAGTGTCGTCGTTCGGGATCGGCGGCACGAATGCGCATGTGATTTTGGAG 2730
Db 6287 CGCCGGGCCGCCGTCTCCTCCTTCGGGATCAGCGGCACCAATGCGCATGTGGTGCTCGAA 6346
Qy 2731 GAGGCGCCGGCGGCGGGGGGCTGTGTTGCCGGGGGTGGGGTGTTGGAGGGTGCTCCGGGT 2790
Db 6347 GAGGCCCCGGTGGTTGTCGAGGGTGCTTCGGTCGTCGAGCCGTCGGTTGGCGGGTCGGCG 6406
Qy 2791 CTTGCCATTTCGGTGGCTGAGTCGGTGGCCGCTCCAGTGGCTGTGTCTGCGCCGGTGGCT 2850
Db 6407 GTCGGCGGCGGTGTGACGCCTTGGGTGGTGTCGGCGAAGTCCGCTGCCGCGCTCGACGCG 6466
Qy 2851 GAGTCGGTGCCGGTGCCGGTGCCGGTGCCGGTTCCTGTGCCGGTGTCGGCTAGGTCTGAG 2910
Db 6467 CAGATCGAGCGGCT----------TGCCGCATTCGCCTCGCGGGATC 6503
Qy 2911 GCTGGGTTGCGGGCGCAGGCGGAGGCGTTGCGTCAGTACGTGGCAGTCCGGCCGGACGTT 2970
Db 6504 GTACGGATGACGCCGACGCCGGTGCTGTCGACGC---------- 6537
Qy 2971 TCGCTTGCCGATGTGGGTGCGGGTCTGGCCTGTGGGCGGGCTGTGCTGGAGCATCGTGCG 3030
Db 6538 -----GGGCGCTGTCGCTCACGTACTGGCTGACGGGCGTGCTCAGTTCGAGCACCGGGCC 6592
Qy 3031 GTCGTCCTGGCCGCGGACCGTGAGGAGCTGGTGCAAGGGTTGGGGGCGCTGGCGGCGGGT 3090
Db 6593 GTCGCGCTCGGCGCCGGGGCGGACGACCTCGTACAGGCGCTGGCCGATCCGGACGGG--- 6649
Qy 3091 GAGCCGGATCGGCGGGTGACCACGGGTCATGCGCCGGGTGGTGACCGGGGCGGTGTCGTC 3150
Db 6650 ----------CTGATACGCGGAACGGCTTCCGGTGTCGGGCGAGTGGCG 6688
Qy 3151 TTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCC 3210
Db 6689 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCTGGCATGGGTGCCGAACTGCTGGACTCT 6748
Qy 3211 TCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGAC 3270
Db 6749 TCCGCGGTGTTCGCGGCGGCCATGGCCGAGTGTGAGGCCGCGCTGTCCCCGTACGTCGAC 6808
Qy 3271 TGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGAT 3330
Db 6809 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 6868
Qy 3331 GTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTAC 3390
Db 6869 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 6928
Qy 3391 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 3450
Db 6929 GGTGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 6988
Qy 3451 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 3510
Db 6989 GCCGGAGCCCTGCCCCTGGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 7048
Qy 3511 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 3567
Db 7049 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAACGAGGACGCCGTC 7108
Qy 3568 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 3627
Db 7109 CTGGAGCGACTGAGTGA------CTTCGACGGGCTGTCCGTCGCCGCCGTCAACGGGCCC 7162
Qy 3628 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 3687
Db 7163 ACCGCCACTGTCGTGTCGGGTGACCCCGTACAGATCGAAGAGCTTGCTCAGGCGTGCAAG 7222
Qy 3688 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 3747

```
Db      16412 CCGCGAACAGGGAGGCCAGGAGCGTCTGGTCACCTCACTCGCCGAAGCCTGGACCAACGG 16471

Qy      14721 GCACCCCCACCACTACACCCACCACCACAACCAACCCCACACCCACACCCACCTCGACCT 14780

Db      16472 CCTCACCATCGACTGGGCGCCCGTCC---TCCCCACCGCAACCGGCCACCACCCCGAGCT 16528

Qy      14781 CCCCACCTACCCCTTCCAACACCACCACTACTGGCTCCA 14819

Db      16529 CCCCACCTACGCCTTCCAGCGCCGTCACTACTGGCTCCA 16567


RESULT 2
US-09-860-846-5
; Sequence 5, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT:  Liu, H.
;  APPLICANT:  Xue, Y.
;  APPLICANT:  Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/860,846
;  CURRENT FILING DATE:  2001-05-18
;  PRIOR APPLICATION NUMBER: 09/105,537
;  PRIOR FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 5
;   LENGTH: 36778
;   TYPE: DNA
;   ORGANISM: Streptomyces venezuelae
US-09-860-846-5

  Query Match    .          9.6%;  Score 2951.8;  DB 9;  Length 36778;
  Best Local Similarity   54.4%;  Pred. No. 0;
  Matches 8436;  Conservative    0;  Mismatches 5722;  Indels 1361;  Gaps   69;

Qy          8 GGATGGACGGCGGGGAAGAACCCCGCCCTGCGGCAGGGGAGGTCCTCGGAGTGGCCGACG 67

Db       3561 GGCCGGACGATCTGGCGGCAGCGCTGGCCGCGCCTGAGGGTCTGGTCCGGGGCGTGGCTT 3620

Qy         68 AGGCGGACGGCGGCGTCGTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGG 127

Db       3621 CCGGTGTCGGGCGAGTGGCGTTCGTGTTCCCCGGGCAGGGCACGCAGTGGGCCGGCATGG 3680

Qy        128 GAAGGGAACTTCTCGACGCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCG 187

Db       3681 GTGCCGAACTGCTGGACTCTTCCGCGGTGTTCGCGGCGGCCATGGCCGAATGCGAGGCCG 3740

Qy        188 CGTTCGCGCCCTACGTCGACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTC 247

Db       3741 CACTCTCCCCGTACGTCGACTGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGC 3800

Qy        248 CCGGGCTGGACCGGGTGGACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGG 307

Db       3801 CCACGCTGGAGCGGGTCGATGTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGG 3860

Qy        308 CCGCCCTCTGGCGCTCGCAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCG 367

Db       3861 CTCGCGTGTGGCAGCACCACGGGGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCG 3920

Qy        368 AGATCGCGGCAGCCCACGTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGA 427

Db       3921 AGATCGCCGCCGCGTACGTCGCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGA 3980

Qy        428 CGCTTTGGAGCCAGGCACAGACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCG 484

Db       3981 CCCTGCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCG 4040

Qy        485 CCGCCACGCCGGATGAGCTCCTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGC 544

Db       4041 CGCTGAGCGAGGACGCCGTCCTGGAGCGACTGGCCGGGTTCGACG--------------- 4085

Qy        545 GGCTCGCCGTCGCAGCCGTCAACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGG 604

Db       4086 GGCTGTCCGTCGCCGCTGTGAACGGGCCCACCGCCACCGTGGTCTCCGGTGACCCCGTAC 4145

Qy        605 CCGTCGCGGACCTGGTGGCCGACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGG 664

Db       4146 AGATCGAAGAGCTTGCTCGGGCGTGTGAGGCCGATGGGGTCCGTGCGCGGGTCATTCCCG 4205

Qy        665 TGGACGTTCCCGCCCACTCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCC 724

Db       4206 TCGACTACGCGTCCCACAGCCGGCAGGTCGAGATCATCGAGAGCGAGCTCGCCGAGGTCC 4265

Qy        725 TGCTGCCCATCACCCCACGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCC 784

Db       4266 TCGCCGGGCTCAGCCCGCAGGCTCCGCGCGTGCCGTTCTTCTCGACACTCGAAGGCGCCT 4325

Qy        785 GCCTCGACACCCGCGAGCTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCC 844

Db       4326 GGATCACCGAGCCCGTGCTCGACGGCGGCTACTGGTACCGCAACCTGCGCCATCGTGTGG 4385

Qy        845 GGTTCGAGCCCGCCGCCCGG---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGA 901

Db       4386 GCTTCGCCCCGGCCGTCGAGACCCTGGCCACCGACGAGGGCTTCACCCACTTCGTCGAGG 4445

Qy        902 TGAGCCCGCACCCGGTGCTGACCATGGGCCTCCAGGAG---------------------- 939

Db       4446 TCAGCGCCCACCCCGTCCTCACCATGGCCCTCCCCGGGACCGTCACCGGTCTGGCGACCC 4505

Qy        940 ----------------------------CTCGCCCCGGACCTGGGCGACACCACCGGCA 970

Db       4506 TGCGTCGCGACAACGGCGGTCAGGACCGCCTAGTCGCCTCCCTCGCCGAAGCATGGGCCA 4565

Qy        971 CCGCCGACACCGTGATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCC 1030

Db       4566 ACGGACTCGCGGTCGACTGGAGCCCGCTCCTCCCCTCCGCGACCGGCCACCACTCCGACC 4625

Qy       1031 TGACGTCTCTCGC---------------CCAACTACGGGGGCATGGTGAGACGTCGGCGA 1075

Db       4626 TCCCCACCTACGCGTTCCAGACCGAGCGCCACTGGCTGGGCGAGATCGAGGCGCTCGCCC 4685

Qy       1076 CCACCGTCCTCTCGGCACGCCTGACCGCGCTGTCCCCCACGCAG---------------- 1119

Db       4686 CGGCGGGCGAGCCGGCGGTGCAGCCCGCCGTCCTCCGCACGGAGGCGGCCGAGCCGGCGG 4745

Qy       1120 -----CAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGCGGTGCTGAACG 1174

Db       4746 AGCTCGACCGGGACGAGCAGCTGCGCGTGATCCTGGACAAGGTCCGGGCGCAGACGGCCC 4805

Qy       1175 ACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTTCGCCCACCTCG 1234

Db       4806 AGGTGCTGGGGTACGCGACAGGCGGGCAGATCGAGGTCGACCGGACCTTCCGTGAGGCCG 4865

Qy       1235 GCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGC 1294

Db       4866 GTTGCACCTCCCTGACCGGCGTGGACCTGCGCAACCGGATCAACGCCGCCTTCGGCGTAC 4925

Qy       1295 GGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGCCGCGCGCCTTC 1354

Db       4926 GGATGGCGCCGTCCATGATCTTCGACTTCCCCACCCCCGAGGCTCTCGCGGAGCAGCTGC 4985

Qy       1355 GGACCGCGGCGCTCGGCCACCTCGACGAGGACACCGCGCCCGTACCGGACTCACCCAGCG 1414

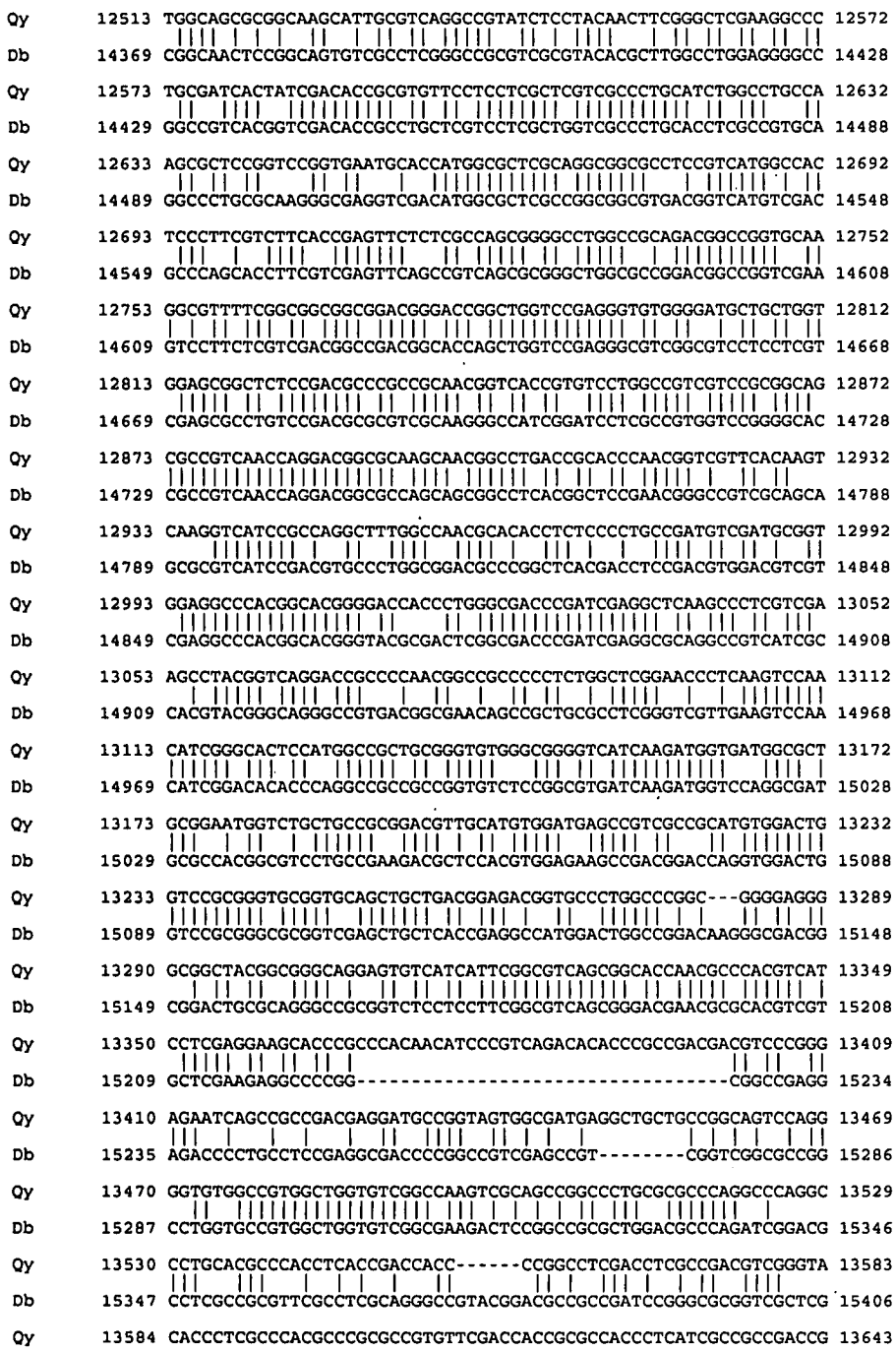
Db       4986 TCCTCGTCGTGCACGG-------GGAGGCGGCGGCGAACCCGGCCGGTGCGGAGCCGGCT 5038

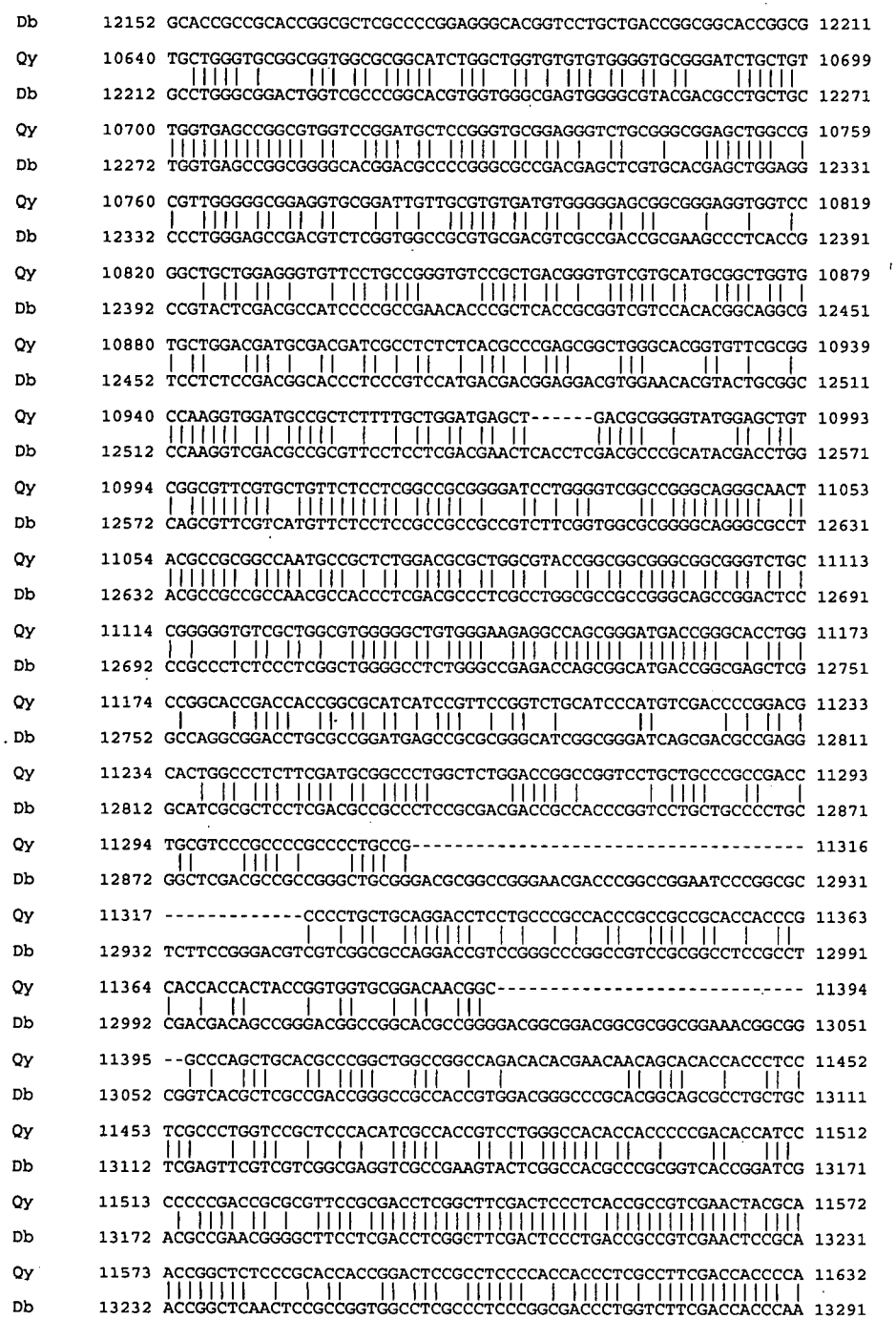
Qy       1415 GCCACGGAGGC-ACGGCAGCGGCGGACGACCCGATCGCCATCATCGGCATGGCATGCCGT 1473

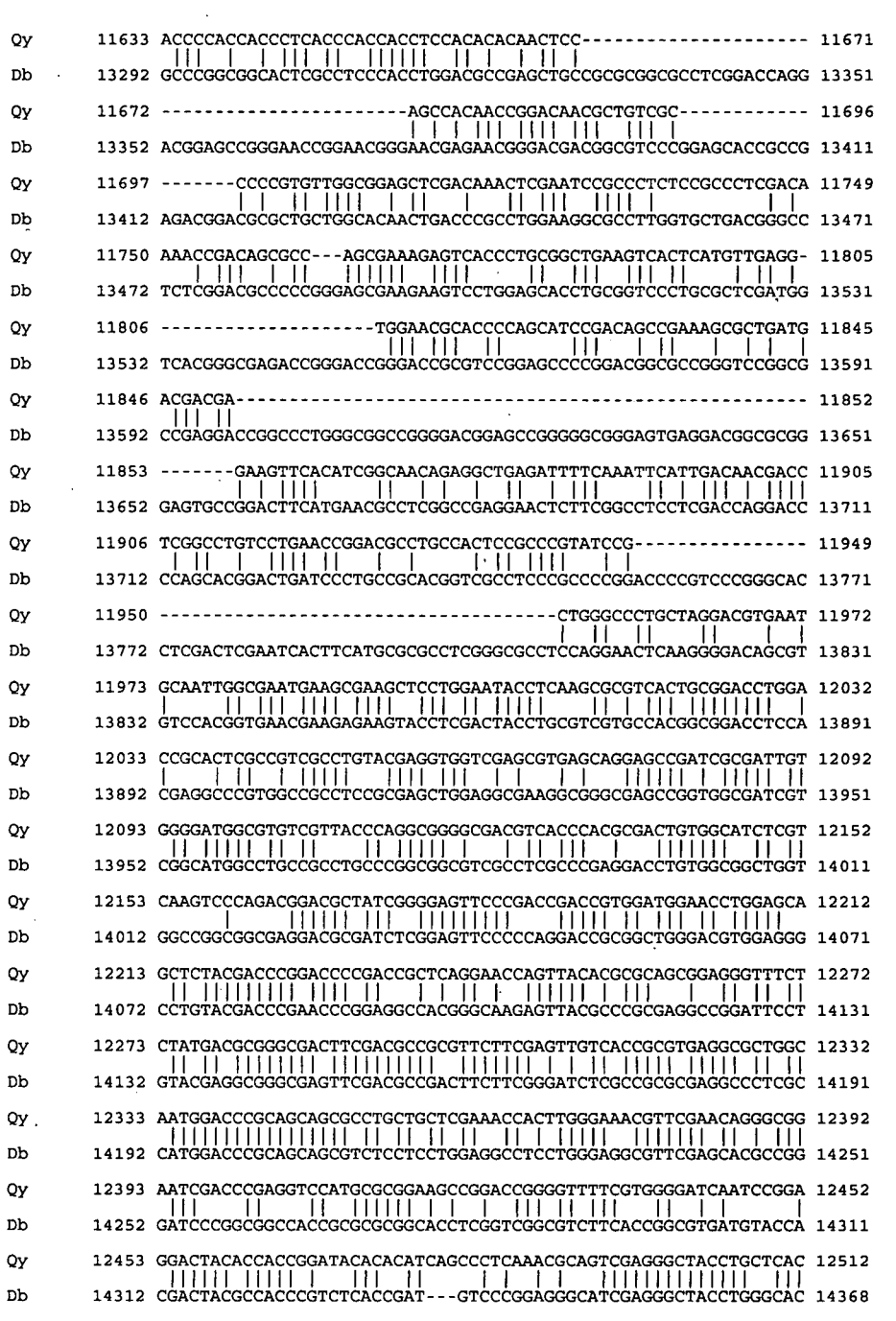
Db       5039 CCGGTGGCGGCGGCCGGTGCCGTCGACGAGCCGGTGGCGATCGTCGGCATGGCCTGCCGC 5098

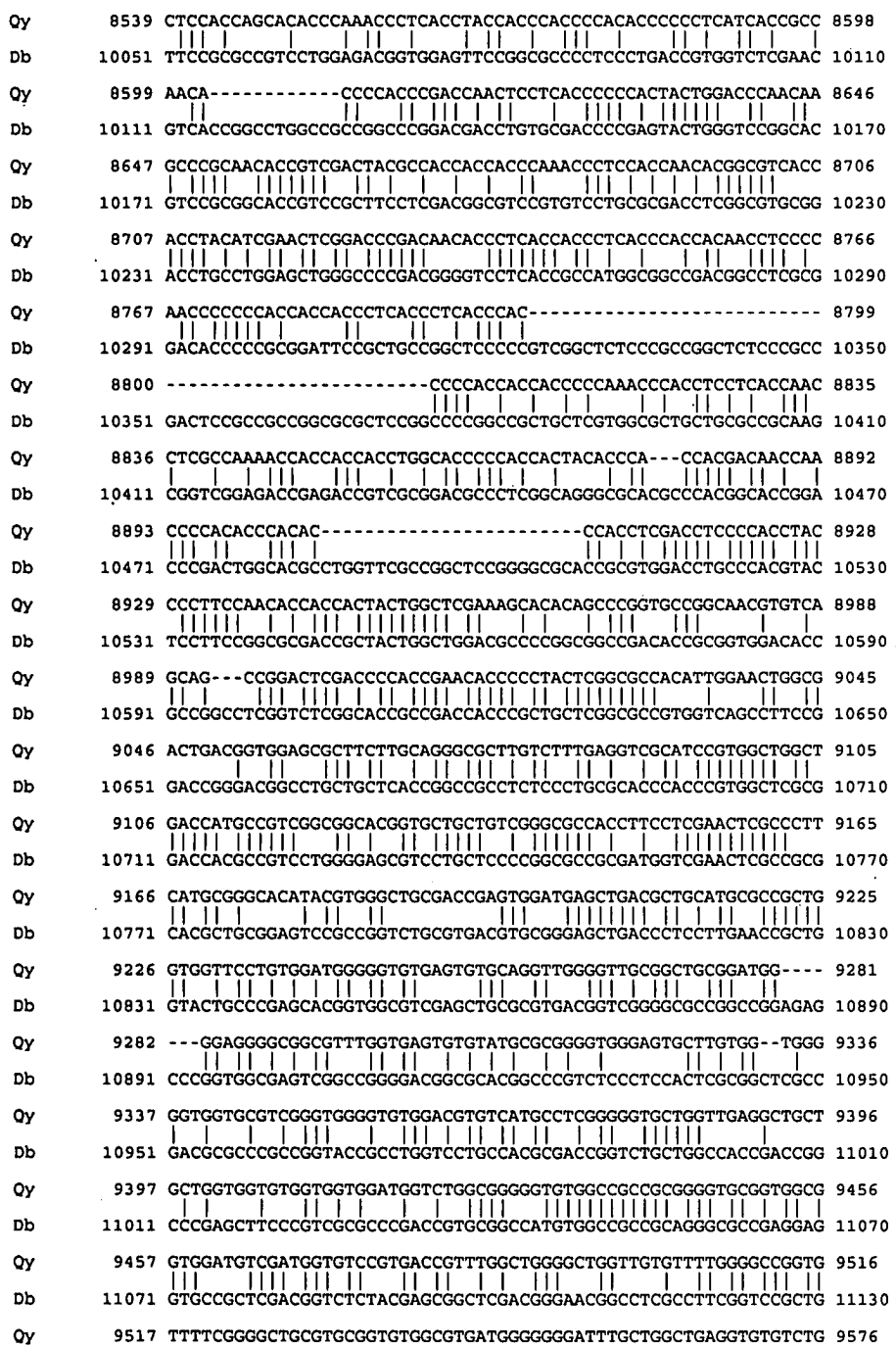
Qy       1474 TTCCCGGGCGGAGTCCGGTCCCCGAAGGACCTGTGGGAGCTGGCCGCCTCGGGCGGAGAC 1533

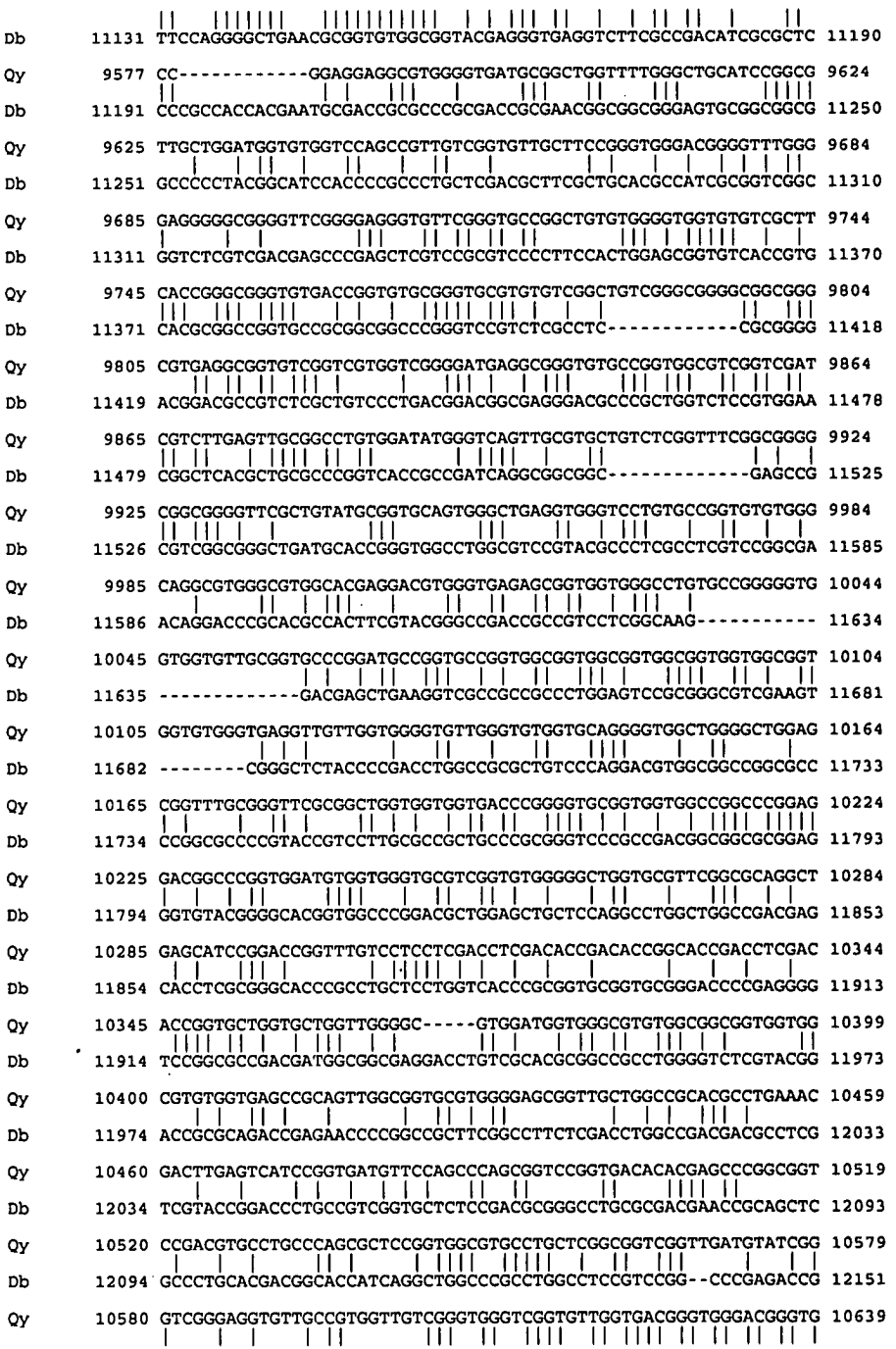
Db       5099 CTGCCCGGTGGGGTCGCCTCGCCGGAGGACCTGTGGCGGCTGGTGGCCGGCGGCGGGGAC 5158
```

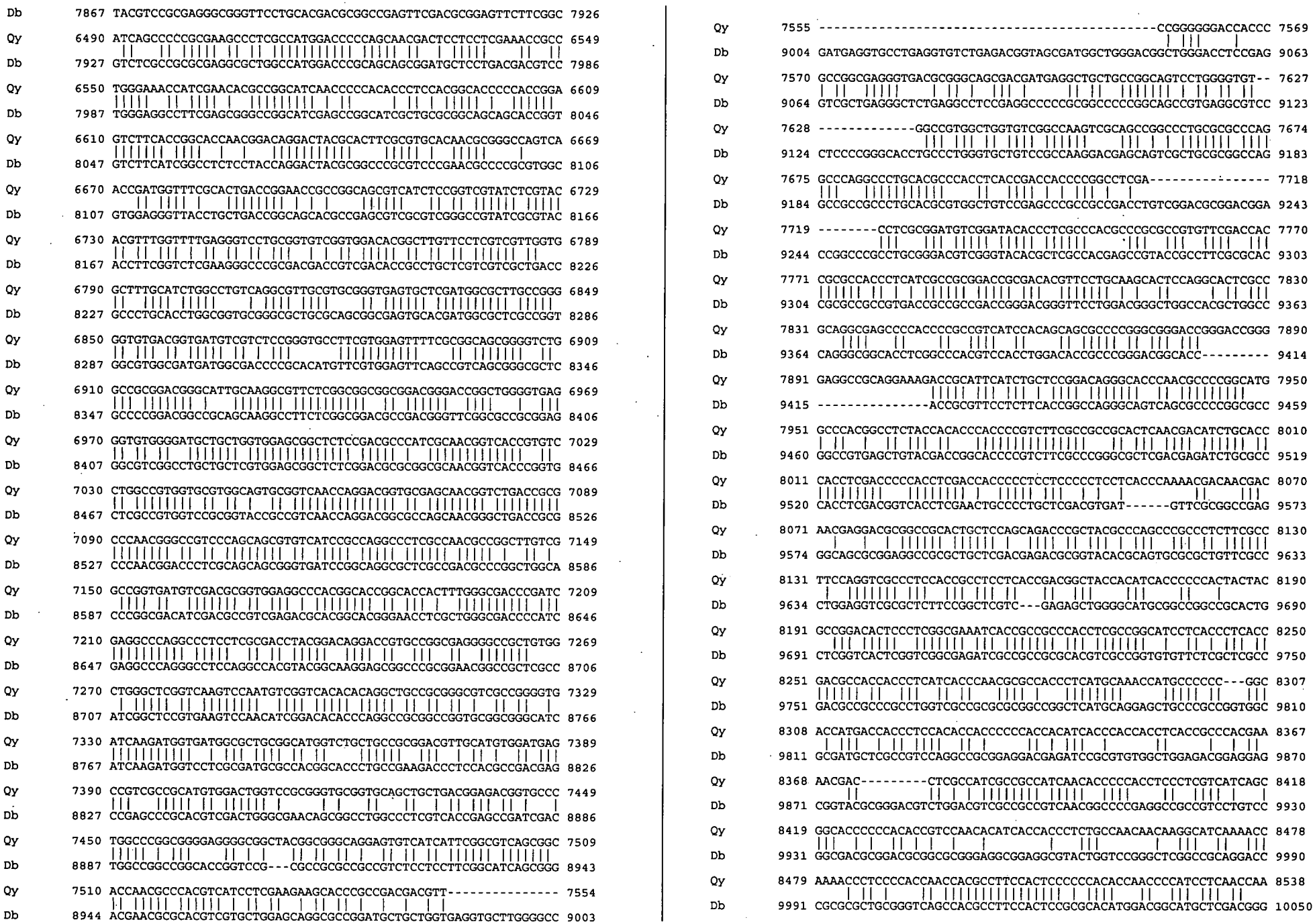

Qy 12513 TGGCAGCGCGGCAAGCATTGCGTCAGGCCGTATCTCCTACAACTTCGGGCTCGAAGGCCC 12572
Db 14369 CGGCAACTCCGGCAGTGTCGCCTCGGGCCGCGTCGCGTACACGCTTGGCCTGGAGGGGCC 14428
Qy 12573 TGCGATCACTATCGACACCGCGTGTTCCTCCTCGCTCGTCGCCCTGCATCTGGCCTGCCA 12632
Db 14429 GGCCGTCACGGTCGACACCGCCTGCTCGTCCTCGCTGGTCGCCCTGCACCTCGCCGTGCA 14488
Qy 12633 AGCGCTCCGGTCCGGTGAATGCACCATGGCGCTCGCAGGCGGCGCCTCCGTCATGGCCAC 12692
Db 14489 GGCCCTGCGCAAGGGCGAGGTCGACATGGCGCTCGCCGGCGGCGTGACGGTCATGTCGAC 14548
Qy 12693 TCCCTTCGTCTTCACCGAGTTCTCTCGCCAGCGGGGCCTGGCCGCAGACGGCCGGTGCAA 12752
Db 14549 GCCCAGCACCTTCGTCGAGTTCAGCCGTCAGCGCGGGCTGGCGCCGGACGGCCGGTCGAA 14608
Qy 12753 GGCGTTTTCGGCGGCGGCGGACGGGACCGGCTGGTCCGAGGGTGTGGGGATGCTGCTGGT 12812
Db 14609 GTCCTTCTCGTCGACGGCCGACGGCACCAGCTGGTCCGAGGGCGTCGGCGTCCTCCTCGT 14668
Qy 12813 GGAGCGGCTCTCCGACGCCCGCCGCAACGGTCACCGTGTCCTGGCCGTCGTCCGCGGCAG 12872
Db 14669 CGAGCGCCTGTCCGACGCGCGTCGCAAGGGCCATCGGATCCTCGCCGTGGTCCGGGGCAC 14728
Qy 12873 CGCCGTCAACCAGGACGGCGCAAGCAACGGCCTGACCGCACCCAACGGTCGTTCACAAGT 12932
Db 14729 CGCCGTCAACCAGGACGGCGCCAGCAGCGGCCTCACGGCTCCGAACGGGCCGTCGCAGCA 14788
Qy 12933 CAAGGTCATCCGCCAGGCTTTGGCCAACGCACACCTCTCCCCTGCCGATGTCGATGCGGT 12992
Db 14789 GCGCGTCATCCGACGTGCCCTGGCGGACGCCCGGCTCACGACCTCCGACGTGGACGTCGT 14848
Qy 12993 GGAGGCCCACGGCACGGGGACCACCCTGGGCGACCCGATCGAGGCTCAAGCCCTCGTCGA 13052
Db 14849 CGAGGCCCACGGCACGGGTACGCGACTCGGCGACCCGATCGAGGCGCAGGCCGTCATCGC 14908
Qy 13053 AGCCTACGGTCAGGACCGCCCCAACGGCCGCCCCCTCTGGCTCGGAACCCTCAAGTCCAA 13112
Db 14909 CACGTACGGGCAGGGCCGTGACGGCGAACAGCCGCTGCGCCTCGGGTCGTTGAAGTCCAA 14968
Qy 13113 CATCGGGCACTCCATGGCCGCTGCGGGTGTGGGCGGGGTCATCAAGATGGTGATGGCGCT 13172
Db 14969 CATCGGACACACCCAGGCCGCCGCCGGTGTCTCCGGCGTGATCAAGATGGTCCAGGCGAT 15028
Qy 13173 GCGGAATGGTCTGCTGCCGCGGACGTTGCATGTGGATGAGCCGTCGCCGCATGTGGACTG 13232
Db 15029 GCGCCACGGCGTCCTGCCGAAGACGCTCCACGTGGAGAAGCCGACGGACCAGGTGGACTG 15088
Qy 13233 GTCCGCGGGTGCGGTGCAGCTGCTGACGGAGACGGTGCCCTGGCCCGGC---GGGGAGGG 13289
Db 15089 GTCCGCGGGCGCGGTCGAGCTGCTCACCGAGGCCATGGACTGGCCGGACAAGGGCGACGG 15148
Qy 13290 GCGGCTACGGCGGGCAGGAGTGTCATCATTCGGCGTCAGCGGCACCAACGCCCACGTCAT 13349
Db 15149 CGGACTGCGCAGGGCCGCGGTCTCCTCCTTCGGCGTCAGCGGGACGAACGCGCACGTCGT 15208
Qy 13350 CCTCGAGGAAGCACCCGCCCACAACATCCCGTCAGACACACCCGCCGACGACGTCCCGGG 13409
Db 15209 GCTCGAAGAGGCCCCGG----------CGGCCGAGG 15234
Qy 13410 AGAATCAGCCGCCGACGAGGATGCCGGTAGTGGCGATGAGGCTGCTGCCGGCAGTCCAGG 13469
Db 15235 AGACCCCTGCCTCCGAGGCGACCCCGGCCGTCGAGCCGT--------CGGTCGGCGCCGG 15286
Qy 13470 GGTGTGGCCGTGGCTGGTGTCGGCCAAGTCGCAGCCGGCCCTGCGCGCCCAGGCCCAGGC 13529
Db 15287 CCTGGTGCCGTGGCTGGTGTCGGCGAAGACTCCGGCCGCGCTGGACGCCCAGATCGGACG 15346
Qy 13530 CCTGCACGCCCACCTCACCGACCACC------CCGGCCTCGACCTCGCCGACGTCGGGTA 13583
Db 15347 CCTCGCCGCGTTCGCCTCGCAGGGCCGTACGGACGCCGCCGATCCGGGCGCGGTCGCTCG 15406
Qy 13584 CACCCTCGCCCACGCCCGCGCCGTGTTCGACCACCGCGCCACCCTCATCGCCGCCGACCG 13643
Db 15407 CGTACTGGCCGGCGGGCGCGCCGAGTTCGAGCACCGGGCCGTCGTGCTCGGCACCGGACA 15466
Qy 13644 CGACACCTTCCTGCAAGCACTCCAGGCACTCGCCGCAGGCGAACCCCACCCCGCCGTCAT 13703
Db 15467 GGACGATTTCGCGCAGGCGCTGACCGCTCCGGAAGGACTGATACGCGGCACGCCCTC--- 15523
Qy 13704 CCACAGCAGCGCCCCAGGCGGGACCGGGACCGGGGAGGCCGCAGGAAAGACCGCATTCAT 13763
Db 15524 ----------GGACGTGGGCCGGGTGGCGTTCGT 15547
Qy 13764 CTGCTCCGGACAGGGCACCCAACGCCCCGGCATGGCCCACGGCCTCTACCACACCCACCC 13823
Db 15548 GTTCCCCGGTCAGGGCACGCAGTGGGCCGGGATGGGCGCCGAACTCCTCGACGTGTCGAA 15607
Qy 13824 CGTCTTCGCCGCCGCACTCAACGACATCTGCACCCACCTCGACCCCCACCTCGACCACCC 13883
Db 15608 GGAGTTCGCGGCGGCCATGGCCGAGTGCGAGAGCGCGCTCTCCCGCTATGTCGACTGGTC 15667
Qy 13884 CCTCCTCCCCCTCCTCACCCAGGACCCCAACACCCAGGACACCACCACCCTCGAAGAAGC 13943
Db 15668 GCTGGAGGCCGTCGTCCGGCAGGCGCC----------GGG 15697
Qy 13944 GGCCGCACTGCTCCAGCAGACCCGCTACGCCCAGCCCGCCCTCTTCGCCTTCCAGGTCGC 14003
Db 15698 CGCGCCCACGCTGGAGCGGGTCGACGTCGTCCAGCCCGTGACCTTCGCTGTCATGGTTTC 15757
Qy 14004 CCTCCACCGCCTCCTCACCGACGGCTACCACATCACCCCCCACTACTACGCCGGACACTC 14063
Db 15758 GCT---GGCGAAGGTCTGGCAGCACCACGGCGTGACGCCGCAGGCCGTCGTCGGCCACTC 15814
Qy 14064 CCTCGGCGAAATCACCGCCGCCCACCTCGCCGGCATCCTCACCCTCACCGACGCCACCAC 14123
Db 15815 GCAGGGCGAGATCGCCGCCGCGTACGTCGCCGGTGCCCTCACCCTCGACGACGCCGCCCG 15874
Qy 14124 CCTCATCACCCAACGCGCCACCCTCATGCAAACCATGCCCCCCGGCAC---CATGACCAC 14180
Db 15875 CGTCGTCACCCTGCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAAGGGCGGCATGAT 15934
Qy 14181 CCTCCACACCACCCCCCACCACATCACCCACCACCTCACCGCCCACGAAAACGACCTCGC 14240
Db 15935 CTCCCTCGCCCTCAGCGAGGAAGCCACCCGGCAGCGCATCGAGAACCTCCACGGACTGTC 15994
Qy 14241 CATCGCCGCCATCAACACCCCCACCTCCCTCGTCATCAGCGGCACCCCCCACACCGTCCA 14300
Db 15995 GATCGCCGCCGTCAACGGCCCCACCGCCACCGTGGTTTCGGGCGACCCCACCCAGATCCA 16054
Qy 14301 ACACATCACCACCCTCTGCCAACAACAAGGCATCAAAACCAAAACCCTCCCCACCAACCA 14360
Db 16055 AGAGCTCGCTCAGGCGTGTGAGGCCGACGGGGTCCGCGCACGGATCATCCCCGTCGACTA 16114
Qy 14361 CGCCTTCCACTCCCCCCACACCAACCCCATCCTCAACCAACTCCACCAGCACACCCAAAC 14420
Db 16115 CGCCTCCCACAGCGCCCACGTCGAGACCATCGAGAGCGAACTCGCCGAGGTCCTCGCCGG 16174
Qy 14421 CCTCACCTACCACCCACCCCACACCCCCCTCATCACCGCCAACACCCCACCCGACCAACT 14480
Db 16175 GCTCAGCCCGCGGACACCTGAGGTGCCGTTCTTCTCGACACTCGAAGGCGCCTGGATCAC 16234
Qy 14481 CCTCACCCCCCACTACTGGACCCAACAAGCCCGCAACACCGTCGACTACGCCACCACCAC 14540
Db 16235 CGAGCCGGTGCTCGACGGCACCTACTGGTACCGCAACCTCCGCCACCGCGTCGGCTTCGC 16294
Qy 14541 CCAAACCCTCCACCAACACGGCGTCACCACCTACATCGAACTCGGACCCGACAACACCCT 14600
Db 16295 CCCCGCCGTCGAGACCCTCGCCACCGACGAAGGCTTCACCCACTTCATCGAGGTCAGCGC 16354
Qy 14601 CACCACCCTCACCCACGACAACCTCCCCAACACCCCCACCACCACCCTCACCCTCACCCA 14660
Db 16355 CCACCCCGTCCTCACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGCACCCT---CCG 16411
Qy 14661 CCCCCACCACCACCCCCAAACCCACCTCCTCACCAACCTCGCCAAAACCACCACCACCTG 14720

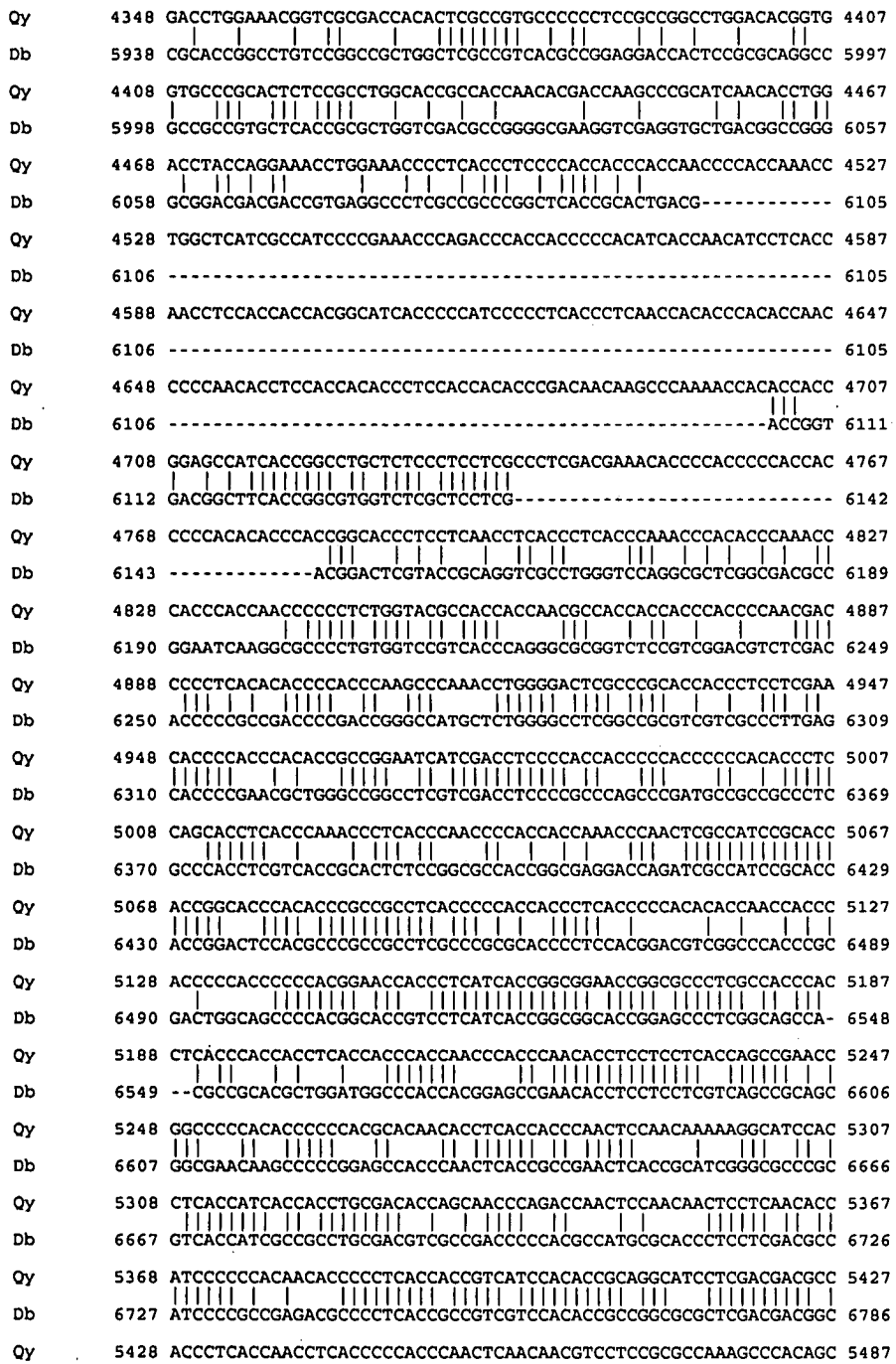


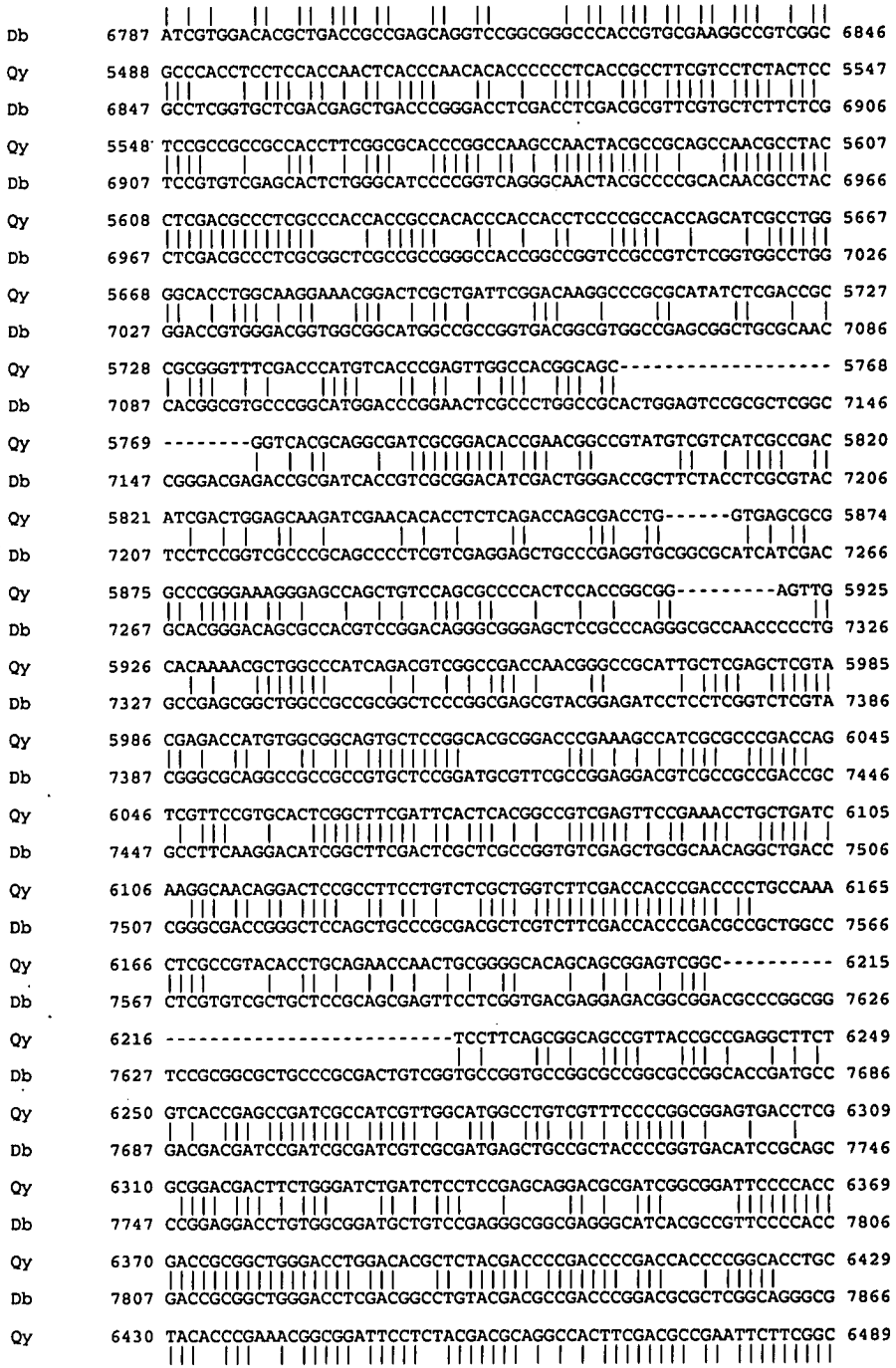


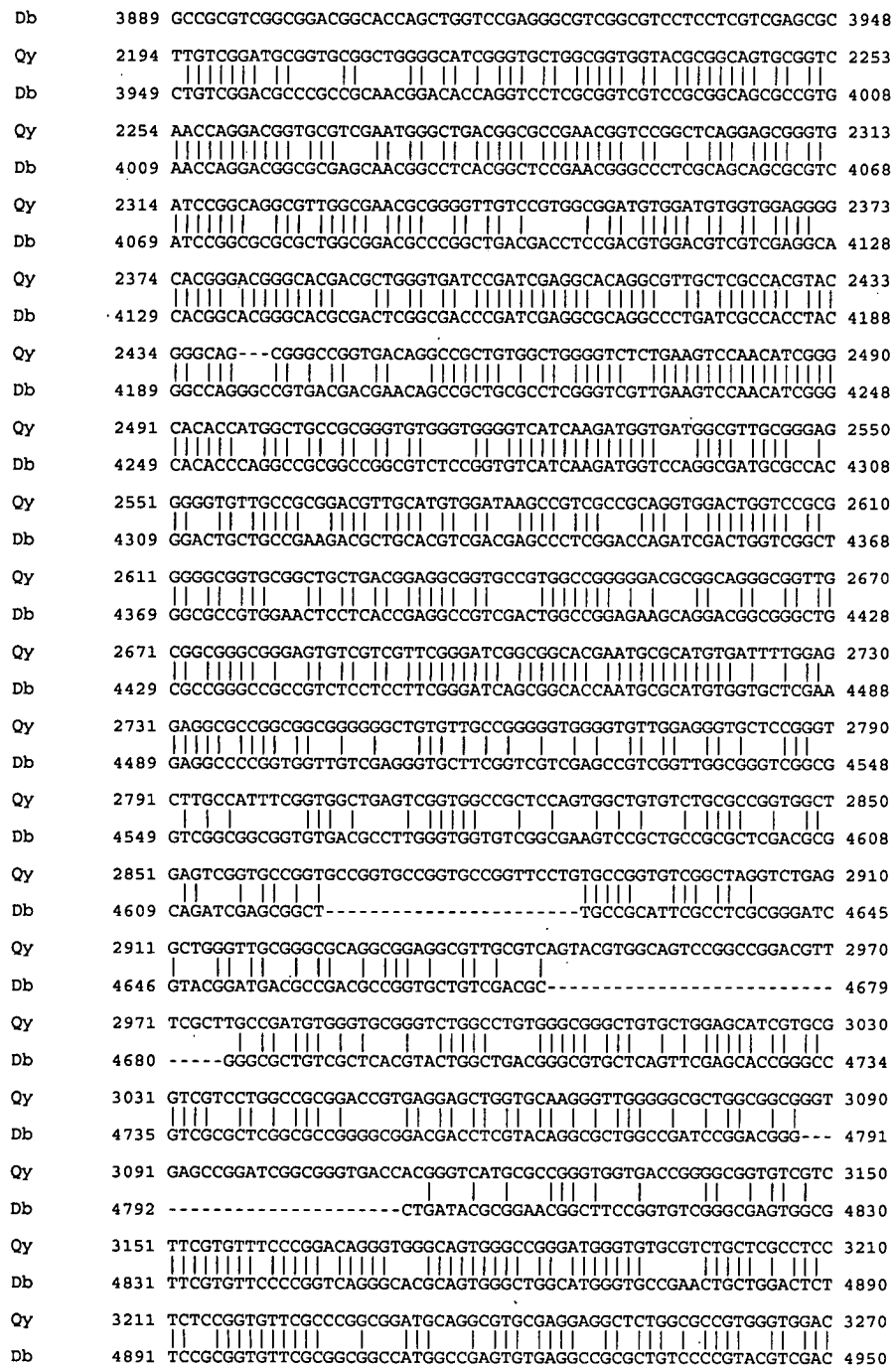


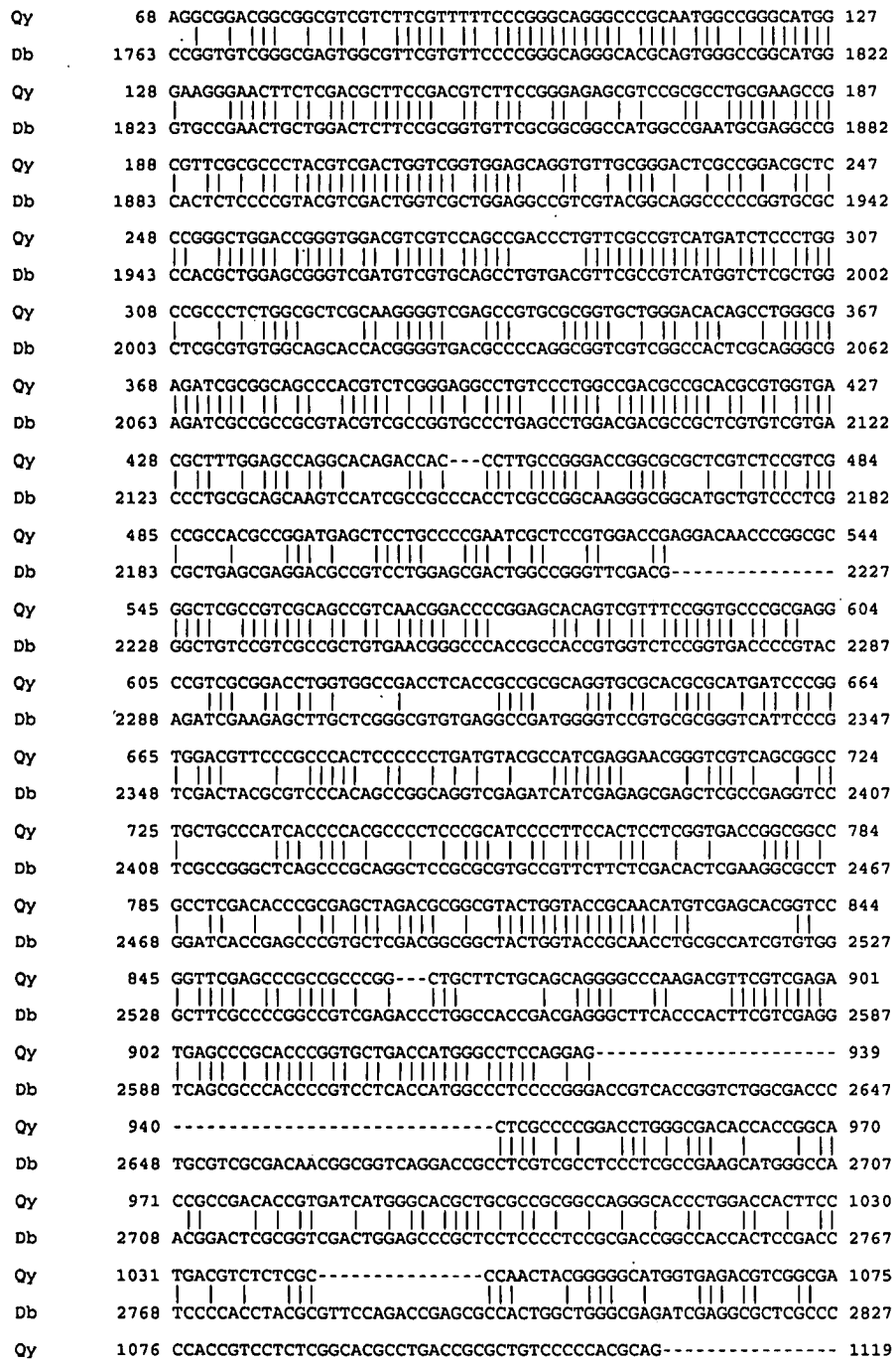









Db 3889 GCCGCGTCGGCGGACGGCACCAGCTGGTCCGAGGGCGTCGGCGTCCTCCTCGTCGAGCGC 3948
Qy 2194 TTGTCGGATGCGGTGCGGCTGGGGCATCGGGTGCTGGCGGTGGTACGCGGCAGTGCGGTC 2253
Db 3949 CTGTCGGACGCCCGCCGCAACGGACACCAGGTCCTCGCGGTCGTCCGCGGCAGCGCCGTG 4008
Qy 2254 AACCAGGACGGTGCGTCGAATGGGCTGACGGCGCCGAACGGTCCGGCTCAGGAGCGGGTG 2313
Db 4009 AACCAGGACGGCGCGAGCAACGGCCTCACGGCTCCGAACGGGCCCTCGCAGCAGCGCGTC 4068
Qy 2314 ATCCGGCAGGCGTTGGCGAACGCGGGGTTGTCCGTGGCGGATGTGGATGTGGTGGAGGGG 2373
Db 4069 ATCCGGCGCGCGCTGGCGGACGCCCGGCTGACGACCTCCGACGTGGACGTCGTCGAGGCA 4128
Qy 2374 CACGGGACGGGCACGACGCTGGGTGATCCGATCGAGGCACAGGCGTTGCTCGCCACGTAC 2433
Db 4129 CACGGCACGGGCACGCGACTCGGCGACCCGATCGAGGCGCAGGCCCTGATCGCCACCTAC 4188
Qy 2434 GGGCAG---CGGGCCGGTGACAGGCCGCTGTGGCTGGGGTCTCTGAAGTCCAACATCGGG 2490
Db 4189 GGCCAGGGCCGTGACGACGAACAGCCGCTGCGCCTCGGGTCGTTGAAGTCCAACATCGGG 4248
Qy 2491 CACACCATGGCTGCCGCGGGTGTGGGTGGGGTCATCAAGATGGTGATGGCGTTGCGGGAG 2550
Db 4249 CACACCCAGGCCGCGGCCGGCGTCTCCGGTGTCATCAAGATGGTCCAGGCGATGCGCCAC 4308
Qy 2551 GGGGTGTTGCCGCGGACGTTGCATGTGGATAAGCCGTCGCCGCAGGTGGACTGGTCCGCG 2610
Db 4309 GGACTGCTGCCGAAGACGCTGCACGTCGACGAGCCCTCGGACCAGATCGACTGGTCGGCT 4368
Qy 2611 GGGGCGGTGCGGCTGCTGACGGAGGCGGTGCCGTGGCCGGGGGACGCGGCAGGGCGGTTG 2670
Db 4369 GGCGCCGTGGAACTCCTCACCGAGGCCGTCGACTGGCCGGAGAAGCAGGACGGCGGGCTG 4428
Qy 2671 CGGCGGGCGGGAGTGTCGTCGTTCGGGATCGGCGGCACGAATGCGCATGTGATTTTGGAG 2730
Db 4429 CGCCGGGCCGCCGTCTCCTCCTTCGGGATCAGCGGCACCAATGCGCATGTGGTGCTCGAA 4488
Qy 2731 GAGGCGCCGGCGGCGGGGGGCTGTGTTGCCGGGGGTGGGGTGTTGGAGGGTGCTCCGGGT 2790
Db 4489 GAGGCCCCGGTGGTTGTCGAGGGTGCTTCGGTCGTCGAGCCGTCGGTTGGCGGGTCGGCG 4548
Qy 2791 CTTGCCATTTCGGTGGCTGAGTCGGTGGCCGCTCCAGTGGCTGTGTCTGCGCCGGTGGCT 2850
Db 4549 GTCGGCGGCGGTGTGACGCCTTGGGTGGTGTCGGCGAAGTCCGCTGCCGCGCTCGACGCG 4608
Qy 2851 GAGTCGGTGCCGGTGCCGGTGCCGGTGCCGGTTCCTGTGCCGGTGTCGGCTAGGTCTGAG 2910
Db 4609 CAGATCGAGCGGCT----------TGCCGCATTCGCCTCGCGGGATC 4645
Qy 2911 GCTGGGTTGCGGGCGCAGGCGGAGGCGTTGCGTCAGTACGTGGCAGTCCGGCCGGACGTT 2970
Db 4646 GTACGGATGACGCCGACGCCGGTGCTGTCGACGC---------- 4679
Qy 2971 TCGCTTGCCGATGTGGGTGCGGGTCTGGCCTGTGGGCGGGCTGTGCTGGAGCATCGTGCG 3030
Db 4680 -----GGGCGCTGTCGCTCACGTACTGGCTGACGGGCGTGCTCAGTTCGAGCACCGGGCC 4734
Qy 3031 GTCGTCCTGGCCGCGGACCGTGAGGAGCTGGTGCAAGGGTTGGGGGCGCTGGCGGCGGGT 3090
Db 4735 GTCGCGCTCGGCGCCGGGGCGGACGACCTCGTACAGGCGCTGGCCGATCCGGACGGG--- 4791
Qy 3091 GAGCCGGATCGGCGGGTGACCACGGGTCATGCGCCGGGTGGTGACCGGGGCGGTGTCGTC 3150
Db 4792 ----------CTGATACGCGGAACGGCTTCCGGTGTCGGGCGAGTGGCG 4830
Qy 3151 TTCGTGTTTCCCGGACAGGGTGGGCAGTGGGCCGGGATGGGTGTGCGTCTGCTCGCCTCC 3210
Db 4831 TTCGTGTTCCCCGGTCAGGGCACGCAGTGGGCTGGCATGGGTGCCGAACTGCTGGACTCT 4890
Qy 3211 TCTCCGGTGTTCGCCCGGCGGATGCAGGCGTGCGAGGAGGCTCTGGCGCCGTGGGTGGAC 3270
Db 4891 TCCGCGGTGTTCGCGGCGGCCATGGCCGAGTGTGAGGCCGCGCTGTCCCCGTACGTCGAC 4950
Qy 3271 TGGTCTGTGGTGGACATCCTGCGCCGGGACGCGGGGGATGCGGTGTGGGAGCGGGCCGAT 3330
Db 4951 TGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGCCCACGCTGGAGCGGGTCGAT 5010
Qy 3331 GTGGTCCAGCCTGTGCTGTTCAGCGTCATGGTGTCTTTGGCTGCTCTGTGGCGTTCCTAC 3390
Db 5011 GTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGGCTCGCGTGTGGCAGCACCAC 5070
Qy 3391 GGTATCGAACCCGACGCGGTCCTTGGCCATTCCCAGGGCGAGATCGCGGCCGCGCATGTG 3450
Db 5071 GGTGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCGAGATCGCCGCCGCGTACGTC 5130
Qy 3451 TGTGGGGCGCTGAGCCTGAAGGACGCGGCGAAGACTGTTGCGCTGCGCAGCCGGGCGCTG 3510
Db 5131 GCCGGAGCCCTGCCCCTGGACGACGCCGCCCGCGTCGTCACCCTGCGCAGCAAGTCCATC 5190
Qy 3511 GCCGC---TGTGCGGGGCCGGGGCGGCATGGCCTCAGTGCCGCTGCCTGCCCAGGAGGTG 3567
Db 5191 GCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCGCGCTGAACGAGGACGCCGTC 5250
Qy 3568 GAGCAGCTCATTGGTGAGCGGTGGGCGGGGCGGTTGTGGGTGGCGGCGGTCAACGGCCCC 3627
Db 5251 CTGGAGCGACTGAGTGA------CTTCGACGGGCTGTCCGTCGCCGCCGTCAACGGGCCC 5304
Qy 3628 CGCTCCACCGCCGTCTCGGGGGATGCCGAGGCGGTGGACGAGGTGCTGGCGTACTGTGCC 3687
Db 5305 ACCGCCACTGTCGTGTCGGGTGACCCCGTACAGATCGAAGAGCTTGCTCAGGCGTGCAAG 5364
Qy 3688 GGCACCGGGGTGCGGGCCCGGCGGATCCCGGTCGACTATGCCTCGCACTGCCCCCATGTG 3747
Db 5365 GCGGACGGATTCCGCGCGCGGATCATTCCCGTCGACTACGCGTCCCACAGCCGGCAGGTC 5424
Qy 3748 CAGCCCCTGCGGGAGGAGTTGCTGGAGCTGCTGGGGGACATCAGCCCGCAGCCGTCCGGC 3807
Db 5425 GAGATCATCGAGAGCGAGCTCGCCCAGGTCCTCGCCGGTCTCAGCCCGCAGGCCCCGCGC 5484
Qy 3808 GTGCCGTTCTTCTCCACGGTGGAGGGCACCTGGCTGGACACCACAACCCTGGACGCCGCC 3867
Db 5485 GTGCCGTTCTTCTCGACGCTCGAAGGCACCTGGATCACCGAGCCCGTCCTCGACGGCACC 5544
Qy 3868 TACTGGTACCGCAACCTGCACCAGCCGGTCCGTTTCAGCGATGCCGTCCAGGCCCTGGCG 3927
Db 5545 TACTGGTACCGCAACCTCCGTCACCGCGTCGGCTTCGCCCCCGCCATCGAGACCCTGGC- 5603
Qy 3928 GATGACGGACACCGCGTCTTCGTCGAAGTCAGCCCCCACCCCACCCTCGTCCCCGCCATC 3987
Db 5604 ----------CGTCGACGAGGGCTTCACGCACTTCGTCGAGGTCAGCGCCCAC 5646
Qy 3988 GAAGACACCACCGAAGACACCGCCGAAGACGTCACCGCGATCGGCAGCCTCCGCCGCGGC 4047
Db 5647 CCCGTCCTCACCATGACCCTCCCCGAGACCGTCACCGGCCTCGGCACCCTCCGTCGCGAA 5706
Qy 4048 GACAACGACACCCGCCGCTTCCTCACCGCCCTCGCCCACACCCATACCACCGGCATCGGC 4107
Db 5707 CAGGGAGGCCAAGAGCGTCTGGTCACCTCGCTCGCCGAGGCGTGGGTCAACGGGCTTCCC 5766
Qy 4108 ACACCCACCACCTGGCACCACCACTACACCCACCACCACACCCACCCCCACCCCCACACG 4167
Db 5767 GTGGCATGGACTTCGCTCCTGCCCGCCACGGCCTCCCGCCC---------- 5807
Qy 4168 CACCTCGACCTGCCCACCTACCCCTTCCAACACCAGCACTACTGGCTCGAGAGCTCACAG 4227
Db 5808 -----CGGTCTGCCCACCTACGCCTTCCAGGCCGAGCGCTACTGGCTCGAGAACACTCCC 5862
Qy 4228 CCGGGTGCCGGATCCGGTTCGGGTGCCGGTGCCGGTTCGGGTGCCGGTTCCGGGCGGGCA 4287
Db 5863 GCCGC----------CCTGGCCACCG 5878
Qy 4288 GGGACTGCGGGCGGGACGGCAGAGGTGGAGTCGCGGTTCTGGGACGCGGTGGCCCGCCAG 4347
Db 5879 GCGACGACTGGCGCTACCGCATCGACTG-GAAGCGCCTCCCGGCCGCCGAGGGGTCCGAG 5937


Qy 68 AGGCGGACGGCGGCGTCGTCTTCGTTTTTCCCGGGCAGGGCCCGCAATGGCCGGGCATGG 127
Db 1763 CCGGTGTCGGGCGAGTGGCGTTCGTGTTCCCCGGGCAGGGCACGCAGTGGGCCGGCATGG 1822
Qy 128 GAAGGGAACTTCTCGACGCTTCCGACGTCTTCCGGGAGAGCGTCCGCGCCTGCGAAGCCG 187
Db 1823 GTGCCGAACTGCTGGACTCTTCCGCGGTGTTCGCGGCGGCCATGGCCGAATGCGAGGCCG 1882
Qy 188 CGTTCGCGCCCTACGTCGACTGGTCGGTGGAGCAGGTGTTGCGGGACTCGCCGGACGCTC 247
Db 1883 CACTCTCCCCGTACGTCGACTGGTCGCTGGAGGCCGTCGTACGGCAGGCCCCCGGTGCGC 1942
Qy 248 CCGGGCTGGACCGGGTGGACGTCGTCCAGCCGACCCTGTTCGCCGTCATGATCTCCCTGG 307
Db 1943 CCACGCTGGAGCGGGTCGATGTCGTGCAGCCTGTGACGTTCGCCGTCATGGTCTCGCTGG 2002
Qy 308 CCGCCCTCTGGCGCTCGCAAGGGGTCGAGCCGTGCGCGGTGCTGGGACACAGCCTGGGCG 367
Db 2003 CTCGCGTGTGGCAGCACCACGGGGTGACGCCCCAGGCGGTCGTCGGCCACTCGCAGGGCG 2062
Qy 368 AGATCGCGGCAGCCCACGTCTCGGGAGGCCTGTCCCTGGCCGACGCCGCACGCGTGGTGA 427
Db 2063 AGATCGCCGCCGCGTACGTCGCCGGTGCCCTGAGCCTGGACGACGCCGCTCGTGTCGTGA 2122
Qy 428 CGCTTTGGAGCCAGGCACAGACCAC---CCTTGCCGGGACCGGCGCGCTCGTCTCCGTCG 484
Db 2123 CCCTGCGCAGCAAGTCCATCGCCGCCCACCTCGCCGGCAAGGGCGGCATGCTGTCCCTCG 2182
Qy 485 CCGCCACGCCGGATGAGCTCCTGCCCCGAATCGCTCCGTGGACCGAGGACAACCCGGCGC 544
Db 2183 CGCTGAGCGAGGACGCCGTCCTGGAGCGACTGGCCGGGTTCGACG---------- 2227
Qy 545 GGCTCGCCGTCGCAGCCGTCAACGGACCCCGGAGCACAGTCGTTTCCGGTGCCCGCGAGG 604
Db 2228 GGCTGTCCGTCGCCGCTGTGAACGGGCCCACCGCCACCGTGGTCTCCGGTGACCCCGTAC 2287
Qy 605 CCGTCGCGGACCTGGTGGCCGACCTCACCGCCGCGCAGGTGCGCACGCGCATGATCCCGG 664
Db 2288 AGATCGAAGAGCTTGCTCGGGCGTGTGAGGCCGATGGGGTCCGTGCGCGGGTCATTCCCG 2347
Qy 665 TGGACGTTCCCGCCCACTCCCCCCTGATGTACGCCATCGAGGAACGGGTCGTCAGCGGCC 724
Db 2348 TCGACTACGCGTCCCACAGCCGGCAGGTCGAGATCATCGAGAGCGAGCTCGCCGAGGTCC 2407
Qy 725 TGCTGCCCATCACCCCACGCCCCTCCCGCATCCCCTTCCACTCCTCGGTGACCGGCGGCC 784
Db 2408 TCGCCGGGCTCAGCCCGCAGGCTCCGCGCGTGCCGTTCTTCTCGACACTCGAAGGCGCCT 2467
Qy 785 GCCTCGACACCCGCGAGCTAGACGCGGCGTACTGGTACCGCAACATGTCGAGCACGGTCC 844
Db 2468 GGATCACCGAGCCCGTGCTCGACGGCGGCTACTGGTACCGCAACCTGCGCCATCGTGTGG 2527
Qy 845 GGTTCGAGCCCGCCGCCCGG---CTGCTTCTGCAGCAGGGGCCCAAGACGTTCGTCGAGA 901
Db 2528 GCTTCGCCCCGGCCGTCGAGACCCTGGCCACCGACGAGGGCTTCACCCACTTCGTCGAGG 2587
Qy 902 TGAGCCCGCACCCGGTGCTGACCATGGGCCTCCAGGAG---------- 939
Db 2588 TCAGCGCCCACCCCGTCCTCACCATGGCCCTCCCCGGGACCGTCACCGGTCTGGCGACCC 2647
Qy 940 ----------CTCGCCCCGGACCTGGGCGACACCACCGGCA 970
Db 2648 TGCGTCGCGACAACGGCGGTCAGGACCGCCTCGTCGCCTCCCTCGCCGAAGCATGGGCCA 2707
Qy 971 CCGCCGACACCGTGATCATGGGCACGCTGCGCCGCGGCCAGGGCACCCTGGACCACTTCC 1030
Db 2708 ACGGACTCGCGGTCGACTGGAGCCCGCTCCTCCCCTCCGCGACCGGCCACCACTCCGACC 2767
Qy 1031 TGACGTCTCTCGC----------CCAACTACGGGGGCATGGTGAGACGTCGGCGA 1075
Db 2768 TCCCCACCTACGCGTTCCAGACCGAGCGCCACTGGCTGGGCGAGATCGAGGCGCTCGCCC 2827
Qy 1076 CCACCGTCCTCTCGGCACGCCTGACCGCGCTGTCCCCCACGCAG---------- 1119
Db 2828 CGGCGGGCGAGCCGGCGGTGCAGCCCGCCGTCCTCCGCACGGAGGCGGCCGAGCCGGCGG 2887
Qy 1120 -----CAGCAGTCGCTGCTCCTGGACCTGGTGCGCGCCCACACCATGGCGGTGCTGAACG 1174
Db 2888 AGCTCGACCGGGACGAGCAGCTGCGCGTGATCCTGGACAAGGTCCGGGCGCAGACGGCCC 2947
Qy 1175 ACGACGGAAACGAGCGCACCGCGTCGGATGCCGGCCCATCGGCGAGTTTCGCCCACCTCG 1234
Db 2948 AGGTGCTGGGGTACGCGACAGGCGGGCAGATCGAGGTCGACCGGACCTTCCGTGAGGCCG 3007
Qy 1235 GCTTCGACTCCGTCATGGGTGTCGAACTGCGCAACCGCCTCAGCAAGGCCACGGGCCTGC 1294
Db 3008 GTTGCACCTCCCTGACCGGCGTGGACCTGCGCAACCGGATCAACGCCGCCTTCGGCGTAC 3067
Qy 1295 GGTTGCCCGTGACGCTCATCTTCGACCACACCACGCCGGCCGCGGTCGCCGCGCGCCTTC 1354
Db 3068 GGATGGCGCCGTCCATGATCTTCGACTTCCCCACCCCCGAGGCTCTCGCGGAGCAGCTGC 3127
Qy 1355 GGACCGCGGCGCTCGGCCACCTCGACGAGGACACCGCGCCCCGTACCGGACTCACCCAGCG 1414
Db 3128 TCCTCGTCGTGCACGG-------GGAGGCGGCGGCGAACCCGGCCGGTGCGGAGCCGGCT 3180
Qy 1415 GCCACGGAGGC-ACGGCAGCGGCGGACGACCCGATCGCCATCATCGGCATGGCATGCCGT 1473
Db 3181 CCGGTGGCGGCGGCCGGTGCCGTCGACGAGCCGGTGGCGATCGTCGGCATGGCCTGCCGC 3240
Qy 1474 TTCCCGGGCGGAGTCCGGTCCCCGAAGGACCTGTGGGAGCTGGCCGCCTCGGGCGGAGAC 1533
Db 3241 CTGCCCGGTGGGGTCGCCTCGCCGGAGGACCTGTGGCGGCTGGTGGCCGGCGGCGGGGAC 3300
Qy 1534 GCCATCGGGCCGTTCCCCACCGACCGCGGATGGCCCACGGAACAGCGTCACGCCCAGGAC 1593
Db 3301 GCGATCTCGGAGTTCCCGCAGGACCGCGGCTGGGACGTGGAGGGGCTGTACCACCCGGAT 3360
Qy 1594 CCCACGCAGCCCGGCACGTTCTATCCGCAGGGAGGCGGGTTCCTTCACGACGCGGCGCAC 1653
Db 3361 CCCGAGCACCCCGGCACGTCGTACGTCCGCCAGGGCGGTTTCATCGAGAACGTCGCCGGC 3420
Qy 1654 TTCGACGCCGGCTTCTTCGGAATCAGTCCACGTGAGGCACTGGCGATGGATCCGCAGCAG 1713
Db 3421 TTCGACGCGGCCTTCTTCGGGATCTCGCCGCGCGAGGCCCTCGCCATGGACCCGCAGCAG 3480
Qy 1714 CGGCTGCTGCTGGAGACGTCCTGGGAGGCGTTCGAGCGGGCGGGAATCGATCCGCTGTCG 1773
Db 3481 CGGCTCCTCCTCGAAACCTCCTGGGAGGCCGTCGAGGACGCCGGGATCGACCCGACCTCC 3540
Qy 1774 GTACGCGGGTCCCGTACGGGCGTCTTCGCGGGCGCCCTCTCCTTCGACTACGGCCCGCGT 1833
Db 3541 CTGCGGGGACGGCAGGTCGGCGTCTTCACTGGGGCGATGACCCACGAGTACGGGCCGAGC 3600
Qy 1834 ATGGACACCGCGTCGTCGGAGGGCGCCGCGGACGTGGAGGGCCACATCCTCACCGGTACC 1893
Db 3601 CT----------GCGGGACGGCGGGGAAGGCCTCGACGGCTACCTGCTGACCGGCAAC 3648
Qy 1894 ACGGGCAGCGTCCTGTCGGGCCGTATCGCCTACAGCTTCGGGCTGGAAGGGCCGGCGATC 1953
Db 3649 ACGGCCAGCGTGATGTCGGGCCGCGTCTCGTACACACTCGGCCTTGAGGGCCCCGCCCTG 3708
Qy 1954 ACCGTGGACACGGGGTGCTCGGCATCGCTCGTGACGCTGCATCTGGCGTGCCAGTCGCTG 2013
Db 3709 ACGGTGGACACGGCCTGCTCGTCGTCGCTGGTCGCCCTGCACCTCGCCGTGCAGGCCCTG 3768
Qy 2014 CGGTCGGGTGAGTGCACGCTCGCGCTGGCCGGCGGCGTCTCGGTCATGTCCACCCTCGGC 2073
Db 3769 CGCAAGGGCGAGGTCGACATGGCGCTCGCCGGCGGCGTGGCCGTGATGCCCACGCCCGGG 3828
Qy 2074 ATGTTCATCGAGTTCTCCCGGCAGCGCGGGCTGTCGGTGGACGGCAGGTGCAAGGCGTAC 2133
Db 3829 ATGTTCGTCGAGTTCAGCCGGCAGCGCGGGCTGGCCGGGGACGGCCGGTCGAAGGCGTTC 3888
Qy 2134 TCGGCTGCAGCCGACGGCACCGGCTGGGGCGAGGGCGTCGGGATGCTGTTGGTGGAGCGG 2193

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 05:09:26 ; Search time 2475 Seconds
(without alignments)
18196.097 Million cell updates/sec

Title: US-09-914-286-1
Perfect score: 30690
Sequence: 1 gtgcagaggatggacggcgg.....acgaaggtgagacagcgtga 30690

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0.
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2956.6	9.6	38506	9	US-09-793-708-19	Sequence 19, Appl
2	2951.8	9.6	36778	9	US-09-860-846-5	Sequence 5, Appli
3	2951.8	9.6	36778	9	US-09-836-821-5	Sequence 5, Appli
4	2951.8	9.6	36778	10	US-09-861-289-5	Sequence 5, Appli
5	2951.8	9.6	37948	9	US-09-988-384B-5	Sequence 5, Appli
6	2342	7.6	13842	9	US-09-860-846-30	Sequence 30, Appl
7	2342	7.6	13842	9	US-09-988-384B-30	Sequence 30, Appl
8	2342	7.6	13842	9	US-09-836-821-30	Sequence 30, Appl
9	2342	7.6	13842	10	US-09-861-289-30	Sequence 30, Appl
10	2304.2	7.5	50937	9	US-09-808-880-1	Sequence 1, Appli
11	2188.4	7.1	15872	9	US-09-860-846-1	Sequence 1, Appli
12	2188.4	7.1	15872	9	US-09-988-384B-1	Sequence 1, Appli
13	2188.4	7.1	15872	9	US-09-836-821-1	Sequence 1, Appli
14	2188.4	7.1	15872	10	US-09-861-289-1	Sequence 1, Appli
15	2140.6	7.0	11220	9	US-09-860-846-32	Sequence 32, Appl
16	2140.6	7.0	11220	9	US-09-988-384B-32	Sequence 32, Appl
17	2140.6	7.0	11220	9	US-09-836-821-32	Sequence 32, Appl
18	2140.6	7.0	11220	10	US-09-861-289-32	Sequence 32, Appl
19	1004.4	3.3	4041	9	US-09-860-846-36	Sequence 36, Appl
20	1004.4	3.3	4041	9	US-09-988-384B-36	Sequence 36, Appl
21	1004.4	3.3	4041	9	US-09-836-821-36	Sequence 36, Appl
22	1004.4	3.3	4041	10	US-09-861-289-36	Sequence 36, Appl
23	992.4	3.2	4689	9	US-09-860-846-34	Sequence 34, Appl
24	992.4	3.2	4689	9	US-09-988-384B-34	Sequence 34, Appl
25	992.4	3.2	4689	9	US-09-836-821-34	Sequence 34, Appl
26	992.4	3.2	4689	10	US-09-861-289-34	Sequence 34, Appl
27	985	3.2	68750	9	US-10-014-717-1	Sequence 1, Appli
28	637	2.1	15720	9	US-10-025-380-1058	Sequence 1058, Ap
29	637	2.1	15720	10	US-09-922-217-1058	Sequence 1058, Ap
30	637	2.1	15720	10	US-09-833-263-1058	Sequence 1058, Ap
31	481	1.6	5484	9	US-09-712-363-115	Sequence 115, App
32	371.6	1.2	1010	9	US-09-735-056-29	Sequence 29, Appl
33	367	1.2	1973	10	US-09-864-761-3471	Sequence 3471, Ap
34	360.8	1.2	4209	9	US-09-712-363-20	Sequence 20, Appl
35	348	1.1	4851	9	US-09-712-363-116	Sequence 116, App
36	336.8	1.1	1635	10	US-09-864-761-20241	Sequence 20241, A
37	314	1.0	1035	9	US-09-735-056-30	Sequence 30, Appl
c 38	301.8	1.0	12733	9	US-10-032-393-47	Sequence 47, Appl
c 39	301.8	1.0	12739	9	US-10-032-393-8	Sequence 8, Appli
40	272.6	0.9	4830	9	US-09-738-626-3166	Sequence 3166, Ap
c 41	272.6	0.9	3309400	9	US-09-738-626-1	Sequence 1, Appli
42	261	0.9	8730	9	US-10-124-800-1	Sequence 1, Appli
43	253	0.8	615	9	US-10-125-815-3	Sequence 3, Appli
44	253	0.8	925	9	US-09-735-056-1	Sequence 1, Appli
c 45	239.2	0.8	1075	10	US-09-864-761-19241	Sequence 19241, A

ALIGNMENTS

RESULT 1
US-09-793-708-19
; Sequence 19, Application US/09793708
; Publication No. US20030104597A1
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: McDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002121
; CURRENT APPLICATION NUMBER: US/09/793,708
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 09/657,440
; PRIOR FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 09/320,878
; PRIOR FILING DATE: 1999-05-27
; PRIOR APPLICATION NUMBER: US 09/141,908
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: US 09/073,538
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/846,247
; PRIOR FILING DATE: 1997-04-30
; PRIOR APPLICATION NUMBER: US 60/134,990
; PRIOR FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-793-708-19

Query Match 9.6%; Score 2956.6; DB 9; Length 38506;
Best Local Similarity 54.4%; Pred. No. 0;
Matches 8439; Conservative 0; Mismatches 5719; Indels 1361; Gaps 69;

```
Qy          8 GGATGGACGGCGGGGAAGAACCCCGCCCTGCGGCAGGGGAGGTCCTCGGAGTGGCCGACG 67
              || ||||| ||  | | | |  | ||| | | ||  | || | | | |
Db       1703 GGCCGGACGATCTGGCGGCAGCGCTGGCCGCGCCTGAGGGTCTGGTCCGGGGCGTGGCTT 1762
```